視聴者の心拍活動を利用した
映像要約方法の研究

Video Digesting Methods

based on Viewers' Heart Activity

2009年6月

早稲田大学大学院 国際情報通信研究科
国際情報通信学専攻
マルチメディアとヒューマンファクタ研究II

豊沢 聡

目 次

第1章　序論

本研究の目的は、映像の視聴経験を反映した視聴者中心の映像要約方法を確立するところにある。視聴経験に基づくというこのアプローチは、映像が何を呈示しているかではなく、視聴を通じて視聴者が何を感受したかに着目しているという点で、映像内の物理的な特徴を信号処理技術によって検出する従来の映像要約技術と異なる。映像の視聴経験には認知的なものと情動的なものが考えられるが、本研究では、前者については映像の理解のしやすさ、後者については映像の持つ印象の度合いに着目する。そして、「理解しやすい」映像区間を集めた要約映像と、「印象的な」映像区間を集めた要約映像をそれぞれ生成する。視聴経験の取得にあたっては、視聴者の生理心理学的反応を利用するという人間工学的なアプローチをとる。

本章ではまず、映像要約技術の背景を示した上で（1.1 節）、本研究が目指す上記 2 種類の要約映像の目的と意義を説明する（1.2 節）。続いて、映像要約の要素技術の分類を通じて本研究の位置付けを示す（1.3 節）。最後に、以上を踏まえ、本研究のアプローチ方法（1.4 節）と本論文の以降の構成を示す（1.5 節）。

1.1　本研究の背景

映像情報量の爆発的な増加に伴い、従来ながらの方法による映像の管理、編集、視聴、検索が困難になってきている。総務省 情報通信政策研究所の 2008 年度のコンテンツ流通量調査によれば、2002 年 – 2006 年の 5 年間の平均で毎年 87.6 万時間もの商用映像メディアが製作されている [75]。これに家庭用ビデオカメラ（2008 年現在の日本での普及率は 41% [57]）、街頭や施設に敷設されたサーベイランスカメラ、コンピュータ生成された映像、更には近年注目を集めているウェアラブルビデオカメラを介して日

常に見聞した経験をすべてビデオに記録・保管するライフログ映像 [4, 19, 50] といった、個人や組織が国内外で生成する市場を形成しない映像を加えると、毎年蓄積されていく映像量は膨大なものとなる。例えば、インターネット動画共有サイトの YouTube には、推定、毎年 2,300 万時間の映像が投稿されているという [28]。

この爆発的増加を支えているのが、ビデオカメラ、ビデオレコーダ、パーソナルコンピュータといった映像機器の普及と映像情報のディジタル化、インターネットやメディアの多チャンネル化に代表される通信技術の高度化、映像圧縮技術の進展、そしてデータストレージの大容量化と低廉化である（H.263 ／ 64 Kbps 品質であれば、70 年分のライフログが 11 TB で収録できる [4]。これは、2008 年 12 月現在で標準的な大容量ディスクである 1.5 TB HDD 8 個分で、価格にすれば 10 万円を切る）。こうした技術が着実に発展していくことを考えると、映像情報量は今後も増加し続け、それに伴い、映像情報を処理していく上での困難はより深刻になると予想される。

こうした問題に対し、計算機によるキーワードなどのメタデータの付与（インデックシング）、類似性や映像構造の判断、意味理解、編集・検索場面で直感的な操作が可能なユーザインタフェースといったコンテンツ志向のイメージ検索技術（content-based image retrieval）の研究が国際的にも注目を集めている [74]。例えば、日本では 2007 年度から経済産業省の主導による情報大航海プロジェクト [25, 40] が、国際的には 2001 年から米国 国立標準技術研究所（NIST: National Institute of Standards and Technology）が主催する TRECVID（Text Retrieval Conference – Video Retrieval Evaluation [84]）が、産学官協同で正にこうした課題に取り組んでいる。

映像要約（video digesting[1]）はこうした研究の一分野であり、オリジナル映像から代表的な映像区間を自動的に選択・抽出、若しくは不必要と考えられる区間を削減することでオリジナルよりも短い映像を生成する技術を指す [20, 28, 45, 59, 80, 85]。映像要約によって生成された短い映像を、要約映像（video digest）という。どれだけ短くするかは対象となるオリジナル映像や生成される要約映像に求められる内容に応じて異なるが、長ければオリジナルの 50% 前後（教材映像のように、オリジナル映像に

[1] Video abstraction ともいう。

含まれている情報を漏れなく集録する必要のあるタイプの要約映像では、50% 以上の削減で内容把握が困難になるとの報告がある [7])、短いものではビデオクリップのように数% からオリジナル長と無関係に数十秒から数分が想定されている。TRECVID にはラッシュフィルム（撮影時点の未編集映像）の自動編集というタスクがあるが、ここでは全長の 4% に圧縮することが要求されている。

映像要約の主な目的は、オリジナル映像の直感的把握を短時間で可能にするところにある。要約映像は、時間のない視聴者に全編視聴の代替として用いられるだけでなく、全編の一覧表示による概略把握（後述するキーフレーム型)、テキスト検索でキーワードを含んだ文を斜め読みするように検索結果を素早く確認するような場面で活用されることが期待されている。また、映像本編におまけや紹介用のビデオクリップとして添付したり、未編集のまま退蔵されているホームビデオを活用できるように自動編集を施したり、ラッシュフィルムを編集するといった使用法も考えられている。ライフログ映像の場合は、1 日の活動内容を回想するのに 1 日を費やすのは明らかに不合理であり、重要な活動だけを取捨選択した要約映像は必須だといえる。映像データベースの観点からは、例えば高速アクセスが可能な領域に要約映像を、大容量だがアクセス速度が比較的遅い廉価な領域にオリジナル映像をそれぞれ格納することによりシステムの最適化を図るといった利用方法が考えられる。更に、若年層には過剰に刺激的であったり不適切であるような場面をあらかじめ省いておくなど、映像の自動検閲的な役割も期待されている。

映像要約の究極的な目標は、映画の予告編やスポーツ放送のダイジェスト番組のような人手による編集と同等の品質を持つ映像の自動生成にある。しかし、現在の技術では、映像の持つ意味や構造の的確な把握、更にはオリジナル映像には存在しない編集上の意図の付与（e.g. Kerman [41]）は非常に困難である。そこで、現在の主たる研究目標は、所定の性質を持った映像区間を検出することに注がれている。その中でも特に、サッカーゲームのゴールシーン [9]、ニュース番組で特定のキーワードが語られているアナウンサー表示場面 [20]、料理番組で料理手順を示しているショット [51]、といった映像の説明的・客観的な内容の把握に多くが取り組んでいる。こうした研究は

着実に進歩を見せており、例えば、映像要約機能を搭載したビデオレコーダ [47] や要約映像を掲載したインターネットビデオサイト [18, 65] などが一部商用化の途についている。

しかし、何をして映像の代表的な区間とするかの観点は映像のジャンルやユーザの利用目的によって多様であり、現在の映像要約技術がそうした多様性を十分に反映しているとは言いがたい。要約映像をより使いやすく、魅力的なものにしていくには、代表区間の選択方法を幅広く提供していくことが求められる。特に、要約映像の潜在的利用者が「ゴールシーンを集めた映像」や「特定のトピック（キーワード）が登場するニュースのダイジェスト」のように映像の客観的説明を基に検索・視聴したいだけでなく、楽しい番組を観たい、感動する映画を観たいといった、映像を通じて得られる視聴経験に基づいて映像を利用したいことは十分に予期される。つまり、楽しいと感受されるような映像区間やエキサイティングな場面だけを集めることにより、どのような楽しさや興奮がオリジナルに含まれているかを端的に示す要約映像が求められているが、こうしたユーザ中心のアプローチはまだ端緒についたばかりである。

1.2 本研究の目的

上記を背景に、本研究は映像視聴経験に基づいた映像要約方法を確立することにより、要約映像の多様化に貢献することを目的とする。ここで視聴経験とは、映像を視聴することによって視聴者が得る興奮、驚き、喜び、涙をさそう感動といった心情的な感覚や、観た対象を理解するなどの認知的な過程を指す。

映像内の物理的特徴を用いた従来技術の中にも、視聴経験を反映した映像要約がないわけではない。こうした方法では、歓喜を示す歓声、エキサイティングな爆発音、恐怖感があるおどろおどろした音といった特徴を信号処理技術で映像内から抽出し、これらの特徴から視聴者に与える経験を推定することで映像要約が行われる（1.3.3 節で説明）。しかし、こうした映像内の特徴から推測する方法では、例えば映像内で笑いの起こった場面で視聴者が笑うとは限らないように、予測された経験を視聴者が実際に得るとは確実にはいえない。視聴経験に基づくのならば、視聴時の心的状態に基づい

た方がより直接的である。そこで本研究では、映像視聴時の心的状態を反映した生理心理学的反応を利用するという、人間工学的なアプローチをとる。

生理心理学的な反応の収集には生体センサが必要になる。しかし、生体センサの多くはヒトの体にセンサ部分を接触させる必要があり、また配線により体の動きも拘束してしまう。これでは視聴経験自体を損なうだけでなく、データの収集に視聴者の同意が得られないだろう。本研究では、日常的な映像視聴の場面で生理心理学的データを広く収集することを想定しているので、生体センサは無拘束型でなければならない。そこで本研究では、無拘束にデータ取得が可能な心拍活動を利用した方法を検討する。

具体的には、視聴経験の認知的な側面と情動的な側面にそれぞれ注目し、以下の2種類の要約映像の生成方法を確立する。

理解しやすい要約映像

視聴した映像を思い出せないことは多い。そうしたとき、自動的に生成された要約映像を短時間で観返すことができれば、記憶を新たにできる。また、要約映像を第三者にメールなどに添付して送付すれば、どのような映像を観たのかを相手に簡単に示すこともできる。つまり、要約映像にはメモや言葉で書く感想を補強するメッセージの役割を担わせることができる。このような用途では、観る者の視聴時間、ストレージやネットワークの容量といった制約を考えると、要約映像は非常に短いことが望まれる。その反面、極端な要約が必要になるため、理解しにくい内容になることを避けなければならない。

そこで本研究では、理解しやすいと感受された映像区間だけを集めた要約映像の生成を試みる。この目的のため、理解のしやすさは映像視聴における認知的な心的負荷が低い状態であるとして、心的負荷の指標として知られる心拍変動低周波成分を利用する。この映像要約方法は第3章で説明する。

印象的な要約映像

情動に基づく視聴経験型映像要約には、楽しい、感動する、興奮する、リラックスするといった多様な要約上の観点が考えられるが、本研究ではそうした中で映像のもたらす印象の度合いに着目する。そして、印象の度合いが生理心理学的な用語が定義するところの覚醒度と対応するとし、印象的＝覚醒的な映像区間を集めた要約映像の生成方法を確立する。覚醒度の指標には、心拍動と心拍変動高周波成分を利用する。この映像要約方法は第4章で取り扱う。

印象的な要約映像を生成し、サービスとして提供するフレームワークには、Web の既存のレコメンデーションサイトやファンサイトの形式を考えている。こうしたサイトは、対象となる映像を視聴したことのあるユーザが投じるコメントや評点を集計することで、未視聴者が視聴するかどうかなどの判断材料に利用できる平均評点を示すサービスを提供している。本方法が従来と異なるのは、既視聴者から「投票」ではなく、視聴時の心拍活動データを提供してもらう点である。また、評点は映像全体（タイトル）に対して投じられるのではなく、心拍活動を通じて映像の区間単位（e.g. ショット単位）に投じられる。サイトでは、上記2指標を基に区間単位の「票」を集計し、得票から最も印象的とみなされた区間を選択することで要約映像を生成する。この要約映像から、利用者はその映像にどのような印象的な場面があるかを短時間で知ることができるようになる。

1.3 関連研究と問題点

本節では、最初に映像要約の要素技術を分類することで、本研究で生成する要約映像がどの分類に属すかを示す（1.3.1節）。次に、映像要約の一般的な処理手順を説明し、本研究が従来方法のどの部分に貢献するかを明確にする（1.3.2節）。本研究は先に述べたように視聴経験的な観点からの映像要約を目指しているが、これには映像内の物理的特徴量を用いた方法と視聴者の生理心理学的反応を用いた方法がある。続いては、これらの方法の動向と問題点をそれぞれ示すことで本研究の立脚点を示す（1.3.3

節、1.3.4 節)。

1.3.1 映像要約方法の分類

映像要約方法は、1) 要約映像の形式、2) 要約映像に含める映像区間の選択方法、という二つの観点から分類することができる[2]。本研究の映像要約は、1) 動画形式／ハイライト型、2) 映像外情報／生理心理学的反応を用いた区間選択、を志向するものだが、ここではそれぞれの分類の説明をすることで、本研究がこのスタイルを選択した理由を説明する。

形式による分類

形式分類では、まず要約映像が静止画の集合か動画の集合かで大きく分類される (図 1.1)。静止画形式はオリジナル映像を代表すると判断されたフレームを集めることから、キーフレーム (keyframe) 形式とも呼ばれる。その中でも代表フレームを空間的に配置するサムネイル型 (thumbnail) は一覧性が高く、また視聴時に時間の制約を受けないという特徴を持つ。キーフレーム形式はまた、DVD のメニューのようにオリジナル映像への目次インタフェースとしても用いられる。キーフレーム形式の中には、連続したフレームを画像処理によって一つの画像に集約するモザイク型 (mosaic) もある [45, 59]。静止画形式ではオーディオのような時間的なメディアが付随しないのが一般的で、オリジナルからの情報欠落が大きい。また、サムネイル表示された画像は、初見では理解しにくいという指摘もある [7]。カット点検出などの静止画形式の技術の多くは、次に説明する動画形式でも利用されている。

動画形式はオーディオも含む連続したフレームの集合で、要約映像といった場合、一般的にはこちらを指す。オリジナル映像のメディアが保存されるのでキーフレームより情報量が多く、自然な表現形式になっている。

[2] 現状、用語が統一されているとは言いがたいため (特に summary は混用が多い)、本論文では主として Dimitrova と Truong の用語を採用し [20, 85] 、訳語には各種文献を参考に筆者が取捨したものを当てた。

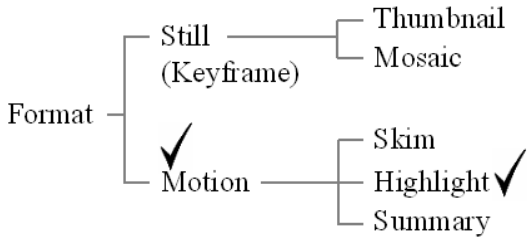

図 1.1: 映像要約の形式（√が本研究のターゲット）

動画形式は更にスキム型（skim）、ハイライト型（highlight）、サマリ型（summary）に分類される。スキム型とは、オリジナル映像全体の印象や理解を損なわずに不要部分を削除、すなわち代表的な区間をできる限り漏らさず選択することで生成する要約映像で、ニュース番組や講義録のような情報志向の映像に適している。ハイライト型はオリジナル映像の中から特定のイベントを抽出して生成する要約映像を指し、スポーツ番組のように、例えば得点シーンを集めることで要約できるタイプの映像に適している。サマリ型はオリジナル映像から構造的又は意味的に重要な部分を抽出するもので、映像構造が重要になるタイプの映像（映画やドラマ）に向く [38, 43, 56]。当然、映像ジャンルの特性に応じてこれらの形式を混ぜて利用することも多い。

本研究では、動画形式をターゲットとする。キーフレーム形式も重要だが、このタイプは選択した映像区間から代表フレームをそれぞれ抽出することでおおむね生成可能である。本方法の生成する要約映像は、動画形式でいえばハイライト型に該当する。これは、印象的な視聴経験をもたらす映像区間はイベント的に出現すると考えられるからである。また、理解しやすい要約映像には非常に短い要約映像が要求されるので、これもハイライト型に相当する。

区間選択方法による分類

映像要約における重要な要素技術は、オリジナル映像の代表的又は不必要な区間を選択する方法である。区間選択方法はその選択の指標となる情報源に応じて、大きく映像内情報型と映像外情報型に分類することができる（図 1.2）。

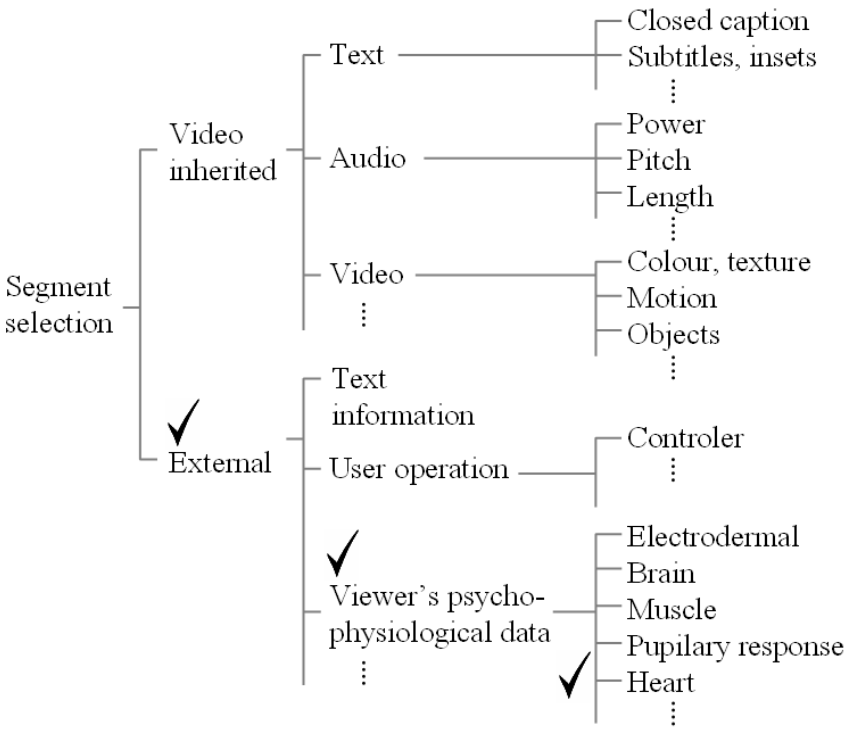

図 1.2: 映像要約における区間選択方法の分類（√が本研究のターゲット）

映像内情報型とは、映像を構成するメディアの信号から物理的特徴量[3]を抽出し、その特徴をオリジナル映像の内容に経験則的に結び付ける方法を指す。例えばテニスゲーム中継番組において、動画から抽出されたコート上の所定の位置に特定の形状の物体（プレーヤ）が存在し、かつその時点でボール打撃音が検出されたらそれはサービス場面であると判定される [52]。オーディオからは音量、ピッチ、時間長など、動画からは色やテクスチャ、動き量、カメラの動き、ボールやコートといったジャンル固有のオブジェクトの位置、形状、数といった特徴が抽出され、それぞれに経験則的に意味付けされる（e.g. フレームを広く緑が占めていればフィールドのロングショット）。また、オーディオ中の発話や画像中のクローズドキャプション、字幕、テロップ、スコア表示などのテキストは抽出されたのち、通常のテキスト処理技術で処理される。更に、カット頻度や台詞・BGM の量やそれらの区切り位置といった、映像の演出的・編集的な要素を抽出して利用する方法もある。映像内情報型は、映像の説明的・客観的な内容を抽出するのに適しており、既存の映像要約技術の大半はこのタイプの方法を用いてい

[3]信号レベル（signal-level）や低レベル（low-level）の特徴とも呼ばれる [24]。

る。映像内情報から視聴経験を推定する試みも幾つかあるが（1.3.3 節で説明）、選択に用いられる物理的指標値と目標とする視聴経験の間接性を考慮すると、次の映像外情報がより適切なアプローチと考えられる。

映像外情報型とは、映像そのもの以外の情報を基に区間選択を行う方法を指す。この方法には、映像作成時のシナリオや別途作成されたトランスクリプト、Web から抽出したキーワード（e.g. オーディオからテキストを抽出するときに、語彙誤りを訂正したり関連語をメタデータとして記録する [26]）などを利用するテキスト処理のアプローチもあるが、映像視聴時の視聴者の行動や心的状態を利用する方法が特に注目を集めている。視聴者の行動に注目した方法には、例えばリモコン操作で早送りした区間は不要区間と判断するといったものがある [48]。しかし、このリモコン方法では、同じ映像若しくは同じ演出スタイルの映像を何度か視聴している視聴者からでなければデータが収集できないため、適用範囲が限られる。心的状態に着目した方法では、視聴者の生理心理学的反応から視聴時の心的状態を推定し、喜んだり感動したりした状態を生起させた映像区間を選択する。この方法には、視聴経験が映像内情報型よりも直接的に反映されるという特徴がある。また、映像内情報型や視聴行動型の方法ほどには映像に依存しないという、方法の適用範囲上の利点もある。この生理心理学的なアプローチによる映像要約については、1.3.4 節で更に説明する。

本研究は、理解しやすい及び印象的であるといった視聴経験は生理心理学的反応を利用することでよりよく検出でき、また方法の適用範囲も広げられると考え、映像外情報型／生理心理学的なアプローチを採用する。生理心理学的な指標には皮膚電気反応、脳活動、筋電、瞳孔やまばたきなどの眼の活動、心拍活動など各種の指標があるが [5, 22, 37]、本研究では、理解や印象と関連付けることができ、また無拘束なデータ取得が可能という他の生理指標にはないメリット（2.1.1 節で説明）を有する心拍活動を用いた映像要約を検討する。

1.3.2 一般的な映像要約手順

本節では、映像要約の一般手順を説明した上で、本研究がターゲットとする領域を示す。映像要約の一般手順は、図 1.3 に示したように、区間分割（segmentation）、区間選択（selection）、区間の短縮化（shortening）、連結（assembly）の 4 ステップで構成される [85]。本研究の目的は、このうちの区間選択の方法を生理心理学的なアプローチにより確立することで、要約映像の多様化に貢献するところにある。

区間分割

最初に、オリジナル映像を要約映像の構成単位となる区間（segment）に分割する。この区間は、有音無音単位といった映像の特性から決定されることもあるが、ショット、シーン、又はシーケンスといった映像の構成要素と一致させるのが一般的である。ここで、ショットは 1 台のカメラから撮影された途切れのないフレーム群を、シーンは一つの場面（同じ撮影場所）でまとめられたショット群を、シーケンスは物語の構成単位でまとめられたシーン群を指す。テキストでいうと、それぞれ単語・文、段落、章節

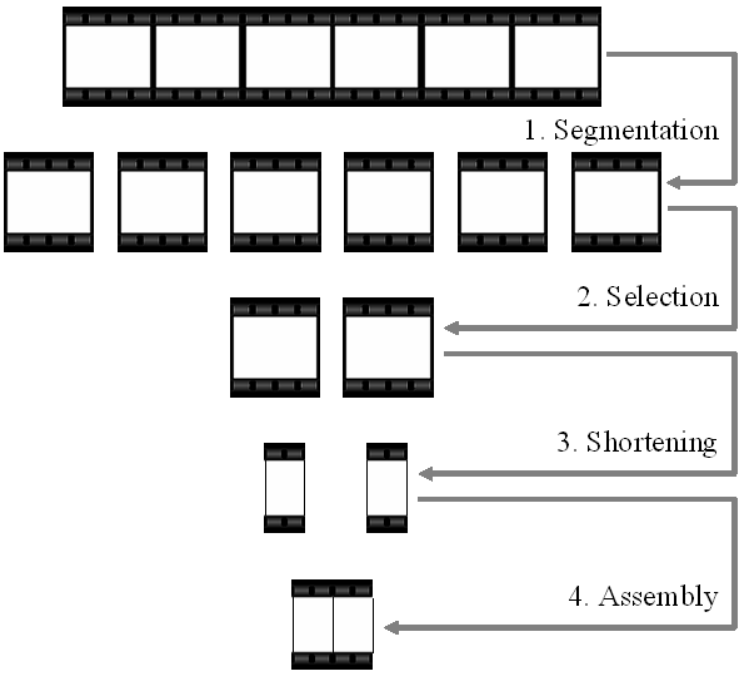

図 1.3: 映像要約の一般手順

に該当する [20]。そのうち、最も用いられるのはショットである。

本研究では、映像区間の単位にショットを用いる。

オリジナル映像をショットに分割するカット点検出（shot boundary detection）技術は、基本的にはカット点（ショットとショットの区切り位置）を挟んだ 2 フレームの物理的特徴の差が、同一ショット内のフレーム間より顕著に異なることを利用している [15, 46]。具体的には、2 フレームの対応するピクセル又はブロックの輝度差を利用する方法、フレーム間の色ヒストグラムの差を統計的手法で検定する方法 [81]、フレームを出入りするエッジ数をカウントする方法、動き量を比較する方法、映像が JPEG/MPEG 形式ならば復号プロセスを経ずに DCT 係数を比較する方法などが提案されている。TRECVID がカット点検出のプロジェクトを終了させたことからもわかるように成熟の進んだ技術であり、現在ではカット点検出はほとんど可能と考えてよい [84, 85]。本研究でも、ショット分割は可能と前提する [4]。

区間選択

映像区間に区間選択の観点と相関している何らかの指標（e.g. 野球ゲームであるならば、区間がヒット場面である確率）を付与することで、オリジナル映像から代表区間を選択、又は不要区間を削除する。区間選択には、先述のように映像内情報を用いる方法と映像外情報を用いる方法がある。いずれにせよ、閾値以上の指標値を持つ区間をすべて選択する（不要区間削除なら閾値以下を省く）、要約映像時間長が定まっているならば指標値の大きい順に選択する、クラスタリングを行い各クラスタの中から 1 区間をそれぞれ選択するといった手順で区間は選択される。

本研究では、理解しやすい要約映像でも印象的な要約映像でも時間長の短いものを生成したいので、あらかじめ定めた区間の数を基に指標値順に選択する方法を採用する。

なお、いずれの区間選択方法を用いるにせよ、オリジナル映像や要約映像へのニーズが多様であることから、一つの方法ですべてのタイプの映像をカバーできるわけで

[4] ホームビデオやライフログ映像のような未編集映像の場合は、有音／無音やカメラワークといった映像の物理的特徴量を用いて適当な長さに区間分割を行う必要があるが、それは本研究の範囲ではない。

はない。また、単一の指標に依存するのではなく、映像内外の情報を含めて複数の指標を組合せて区間選択を行うことが多い。こうしたことから、区間選択方法については、それがどのようなタイプの映像に適切であるかの知見は、精度と同程度に重要である。そこで本研究では、映像を 3 タイプに分類し（2.4 節）、提案方法がどのタイプで効果的であるかも明らかにする。

区間短縮

映像区間は基本的には類似のフレームの集合なので、冗長な部分を省くことでより簡潔にすることができる。区間短縮手順はまた、最終的に生成される要約映像の自然さや見やすさを達成する上で重要である。このような処理を行って得た部分区間を抜粋（excerpt）という。これには、固定時間長を切り出すという機械的な方法、区間内の補助的な指標を用いた方法、無音区間を省く、などの方法が提案されている。

本研究は生理心理学的反応を用いた区間選択方法の確立を主眼としているため、区間選択の妥当性を検証する評価実験に影響する恐れのあるこの手順は基本的には行わない。

但し、理解しやすい要約映像（第 3 章）では、部分的ではあるがこの区間短縮を行っている。これは、最初の区間選択を固定長区間を対象に行っており、要約映像をショット単位で構成するには、補助的な指標を用いて固定長区間からショットを抜粋する必要があるためである。また、評価実験では、ショットや要約映像全体の時間長といった要因が理解しやさに影響することを防ぐため、ショットから固定長の抜粋を更に切り出している。もっとも、これは評価目的であり、本来的に本提案方法がカバーする範囲ではない。

連結

最後に、得られた抜粋（区間短縮を行わない場合は区間）を連結することで要約映像を生成する。この手順は、区間選択や区間短縮がモダリティの連携をあまり考慮せずに行われるため、単純に連結するのではオーディオが途切れたりビデオ部分が急に

切り替わったりするなどの問題が生じることから必要になる。簡単なものでは、カット点で平滑化したりワイプやディゾルブなどのトランジションを挿入する方法がある。また、区間選択が複数のメディアで個別に行われた場合は、メディアが相互に重なり合っている区間を選択する（短め）、若しくはどれか一つのメディアが選択された区間ならば選択する（長め）という手法もある。区間の連結順序はオリジナルの時間順に従わせるのが基本だが、例えば時間順を入れ替えて緩急を巧みに表現するモンタージュ技法的な要素を加味した方法もある。

本研究では、特殊な技法を加えることが区間選択の妥当性の評価実験に影響を与えることを考慮し、区間や抜粋を処理なしでオリジナル時間順に連結している。

1.3.3 映像内情報を用いた経験型映像要約

映像内情報型の映像要約では、1.3.1 節で説明したように説明的内容を観点に区間選択を行うのが基本だが、所定の物理的特徴量と視聴者に与えるであろう視聴経験とを経験則的に関連付けることによって視聴経験を観点とした区間選択を行うことも可能である。代表的な既存研究を表 1.1 に示す。

NTT の chocopara は、音声の強度、高さ、速さから「盛り上がる」区間を選択している [18, 65]。Hanjalic らは、映画とスポーツ番組を対象に、動き量（動きの多い方が覚醒的）、所定の周波数以上のオーディオパワー和（大きい方が覚醒的）、ショット長（ショットが短い方が覚醒的）であると仮定し、これらの特徴量を線形に組合せて情動の動機付けモデル（2.3 節）でいうところの覚醒的な区間を選択している [24]。Arifin らも映画全般を対象に、ショット長、動き量、オーディオのテンポから覚醒度を抽出している [8]。Kang らも同様に映像区間が視聴者に誘起する情動に着目し、映画を対象に、色ヒストグラム、動き情報（強度とカメラワーク）、ショット長といった特徴を HMM（Hidden Markov Model）に投入することで、ショットを恐れ、悲しみ、喜び、平常という 4 種類の情動に分類している [35]。Xu らはオーディオから MFCC （Mel Frequency Cepstrum Coefficients）と音パワーを対象映像の知識領域に結びつけることで、ホラー映画のおどろおどろした音とコメディのサクラの笑いを抽出して恐怖場面

表 1.1: 映像内情報を利用した視聴経験型映像要約方法

	Features	Detected experiences	Target domain †
Chocopara 2005	Audio intensity/pitch /speed	Excitement	Unspecified
Hanjalic 2001	Motion, audio power & shot length	Arousal/Valence	Movie & football game (2)
Arifin 2006	Shot length, motion & audio envelope	Arousal	Movie (4)
Kang 2003	Colour histogram, motion & shot length	Fear, sadness, joy & normal	Movie (6)
Xu 2005	MFCC & audio power	Laughter & horror	Comedy & horror movies (2)
Irie 2008	MFCC	Laughter	Home video (1)
Yoshitaka 2006	Camera movements	Tension, feeling, etc.	Movie (4)

†The number in parentheses shows the number of video types used in their experiments.

と笑いの場面を選択している [86]。同様に、入江らも MFCC を用いて笑いの場面を選択しているが、対象をホームビデオにまで広げている [31]。吉高らはカメラワークには演出上の意図が込められているとし、そこから緊迫感、解放感、心情、孤独感といった感性的な情報を抽出している [89]。

視聴者からデータを収集する必要がないという点から、こうした方法は実用性が高い。しかし、物理的特徴から経験則的に推定される視聴経験が必ずしも実際の経験と一致しない可能性は否定できない。例えば、映像中に笑いが存在しているからといって、そこが実際に笑える場面かはわからない。データ収集が可能ならば、視聴者の直接的な生理心理学的反応を利用した方が確実性は高いといえる。

1.3.4 生理心理学的反応を用いた経験型映像要約

近年、視聴経験、特に情動レベルの経験を反映した映像要約が注目を集めている。その背景には、映像の説明的内容を基にしたこれまでの要約映像だけでは、感性的・経験的な検索や視聴を行いたいというユーザニーズに対応できないという反省がある。視

表 1.2: 生理心理学的反応を利用した視聴経験型映像要約方法

	Measures	Detected experiences	Target Domain †
Healey 1998	GSR	Startle response	Lifelog (1)
Ishijima 2000	EEG (α & β waves)	Excited, attentive & concentrated	Lifelog (1)
Miyata 2004	EEG (β wave)	Deep-thinking	Meeting (1)
Kamada 2001	EMG (Facial)	Laughter	Movie (1)
Kamada 2001	Eye blink	Interested	Movie (1)
Sato 2006	Grip strength	Impression (words)	Movie (3)

†The number in parentheses shows the number of video types used in their experiments.

聴経験は、先述のように映像内情報からも推定可能だが、視聴時の生理心理学的反応に基づいた方がより直接的に取得できる。生理心理学的反応は、人間工学における評価 [37]、ヒトの状態を理解することでより人間的な対応を可能にするロボットやインタフェース（“Affective Computing” と呼ばれる [2, 67]）でも幅広く応用されている。

映像要約の分野では、まず Money らの映像要約（ハイライト型）における生理心理学的データ利用のフレームワークが挙げられるが、得られたデータをどのように指標化するかの詳細は示されていない [55]。同様に、生体センサを埋め込んだビデオカメラにおいて、映像撮影と同時に取得した生理心理学的データから撮影時の感情を算出してメタデータとして埋め込む方法 [49] や、生理心理学的データを後の編集作業に役立てるメカニズム [33, 60] なども提案されてはいるが、これらも生理心理学的データからどのような視聴経験を抽出するかの具体的な用例は提示されていない。これらの事例からも明らかなように、映像要約技術における生理心理学的なアプローチはまだ黎明期にあり、個々のデータをどのように利用するかを検討する段階にあるといえる。

生理心理学的反応の応用例には、表 1.2 のようなものがある。Healey らは指先に取り付けたセンサから得た皮膚電気反応（GSR：Galvanic Skin Response）を利用し、ライフログで記録された映像の中からよく記憶されている場面を選択している [27]。この研究では、驚愕的な状態、すなわち高覚醒な状態になると交感神経活動が賦活し、これに伴い皮膚抵抗が低下するという生理心理学的知見 [5, 37] と、驚愕的な場面はよく記憶

されるという仮説を応用している。石島らは脳波（EEG: Electro-Encephalo-Graphy）を用いて、ライフログ映像から興味関心を惹いた場面を選択している [3, 4, 32]。これは、α 波は瞬間的な興奮や注意、β 波は持続的な興奮や注意でそれぞれ活発化するという知見に基づいている。宮田らも同様に脳波を用い、会議記録映像から会議の重要場面を選択する方法を提案している [53, 54]。鎌田ら及び Shibata らは、顔面の筋電（EMG: Electo-Myo-Graphy）から笑いの表情を検出することで、映画のコミカルな場面を選択している [34, 73]。彼らはまた、興味関心の度合いが瞬目頻度と相関しているという知見 [23] から、興味のある場面を選択している。佐藤らの研究は直接的には映像要約目的ではないが、映像視聴中の握力を感情語対（好き – 嫌い、関心がある – ない、など）と対応付けている [72]。

これらのアプローチは重要であるが、いずれも体部に貼り付けた生体センサと導線で視聴者の体の動きを制約してしまうという問題がある。センサの小型化、無線化は進んではいるものの（e.g. Body Media 社の SenseWear [12]、NeuroSky 社の MindSet [61]、Strauss らの Handwave [76]）、センサを視聴者若しくは撮影者に直接接触させなければならないという問題はつきまとう。ライフログのように各種デバイスを装着するのが日常であるという前提では問題はないであろうが、本研究では、平素に映像視聴を楽しんでいる一般的な視聴者から生理心理学的反応を収集することを前提としているため、拘束型のセンサは不適切である。

その点、心拍活動には無拘束にセンシングが可能な技術が開発されており、本研究の目的にかなっている。取得した心拍活動からは、心拍動そのものと心拍変動の低周波及び高周波成分という三つの指標を抽出できる。このうち、心拍変動低周波成分からは理解しやすい要約映像を、心拍動と心拍変動高周波成分からは印象的な要約映像をそれぞれ生成できる（詳細は第 2 章で説明する）。また、無拘束型心拍センサのほとんどは感圧センサを用いているため、生データから姿勢 [36] や呼吸数 [64, 87] といった指標も抽出できる。心拍活動には、このように一つのセンサから性質の異なる指標を複数取得できるというメリットがあり、コスト面で優れており、指標の種類を拡充していく今後の展開にも有利である。以上を考慮した結果、本研究では視聴経験型の

映像要約に心拍活動を利用することとした。

但し、生理心理学的反応には、視聴者が実際に映像を視聴しないと映像要約に必要なデータが得られないというデメリットがある。しかし、リビングルームのソファやクッション、スポーツバーのスツール、劇場の椅子にセンサを埋め込むことで、視聴者のデータを収集することは原理的に可能である。また、映像編集作業者の椅子にセンサを埋め込むことで、オリジナル映像の作成時のデータを利用することもできる。各所で収集したデータは、ネットワーク経由で映像要約エンジンに転送する。心拍活動データは、RR 間隔時間をテキスト形式で記録しても、視聴 2 時間分で 50 KB 程度にしかならないので、ネットワークやデータベースに負担はかからない。確かに、現時点ではセンサを各家庭に設置するのは現実的ではないだろう。しかし、初期の段階では、例えば試写会の催される劇場といった特定の視聴環境でならば展開可能と考える。

なお、生理心理学的なデータはプライバシーに関わる情報であるため、その収集には十分注意を払わなければならない [67]。本論文ではこの件について考察は行わないが、データ管理ではユーザ情報と関連付けないといった配慮が必要である。

1.4 本研究のアプローチ

本研究は、1.2 節で示したように、理解しやすい及び印象的な要約映像の生成方法を検討する。本章をまとめると、本研究のアプローチ方法は次のようになる。

- 説明的・客観的内容を基にした従来の方法だけでは多様なユーザニーズや映像ジャンルをカバーできないという問題に対し、映像を視聴することにより何が感受されるかという視聴経験に着目した方法を映像要約技術に加えることで、映像要約技術を多様化する。

- 視聴経験の取得には、視聴者の心理生理学的反応を用いるという人間工学的アプローチをとる。映像内の物理的特徴から視聴経験を推定する方法も研究されているが、生理心理学的手法の方がより正確に視聴経験を反映すると考えられるからである。

- 生理心理学的指標には、無拘束なデータ取得が技術的に可能な心拍活動を用いる。心拍活動からは心拍動と心拍変動の低周波及び高周波成分を抽出し、心拍変動低周波成分からは理解しやすさの、心拍動と心拍変動高周波成分からは印象の度合いの指標をそれぞれ抽出する。すなわち、本研究は映像要約技術の中でも、代表的な映像区間をオリジナル映像中から選択する方法の確立に主眼を置く。

- 目的の要約映像の性質に基づき、要約映像の形式は動画／ハイライト型の短いものとする。このとき、映像の最小単位には、確立した検出技術の存在するショットを用いる。ショットから更に抜粋を抽出する区間短縮手順及び区間／抜粋の連結手順については、本研究のスコープ外とする。

1.5 本論文の構成

本論文の構成を以下に示す。

第 1 章 本研究の背景、目的、現在の映像要約技術の動向とその中における本研究の位置付け、本論文の構成を示す。

第 2 章 映像要約に利用する心拍活動指標に関する知見をまとめ、それらの利用方法を説明する。また、以降の章で使用する実験刺激映像と評価基準について説明する。

第 3 章 理解しやすい要約映像の生成方法を検討する。本章では、理解しやすさの定義、理解しやすさと心拍変動低周波成分との関係性と利用方法を説明した上で、映像要約方法を定める。そして、評価実験を通じて、本提案方法の妥当性と適用範囲を検証する。

第 4 章 印象的な要約映像の生成方法を検討する。本章ではまず印象の定義を示し、印象と心拍活動起因の 2 指標（心拍動及び心拍変動高周波成分）との関係性と利用方法を説明する。続いて、これら 2 指標を組合せた映像要約方法を確立するための基礎的知見を得る目的で、評価実験を通じて、2 指標を個別に利用して生成す

る要約映像の性質を明らかにする。その上で、2指標の組合せ方法を示し、評価実験を通じて提案方法の妥当性と適用範囲を示す。

第5章 本研究全体を総括した上で、今後の課題と展望について述べる。

第2章　視聴経験と心拍活動

本章では、「理解しやすい」及び「印象的」であるといった視聴経験に関わる心拍活動の知見をまとめた上で、その処理方法を概説する。

本研究の考え方を模式的に図 2.1 に示す。まず、理解しやすかったり印象的であったりする映像区間を視聴すると、視聴者にはその映像の性質に対応した視聴経験、すなわち精神的な活動が生じる。この精神活動は各種の生理心理学的反応として現れるが、心拍活動の場合は、心拍動の変化やそうした変化をもたらす自律神経系（ANS: Autonomic Nervous System）の変化として外部に現れる。そこで、心拍動及び ANS の変化を測定することにより、呈示された映像区間の性質を推定できると考える。

映像区間の性質を適切に推定するには、理解しやすい及び印象的な映像区間の視聴により生じる精神活動とそれに伴う心拍活動に関する知見と、心拍活動から得られる指標についての知見が必要になる。本章ではまず、心拍活動指標に関する知見と本研究の前提となっている無拘束型心拍センサの動向についてまとめる（2.1 節）。続いて、理解しやすい映像及び印象的な映像の視聴によって生じる反応とその処理方法をそれぞれ説明する（2.2 節、2.3 節）。最後に、第 3 章以降の評価実験で使用する刺激映像と

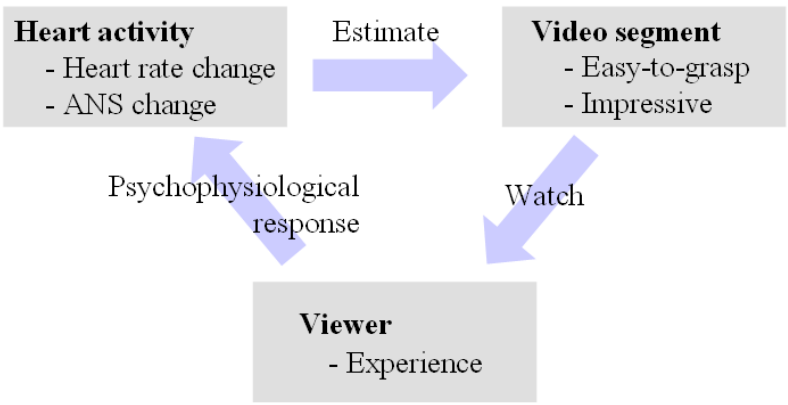

図 2.1: 映像区間 – 視聴経験 – 心拍活動の関係

(2.4 節)、評価に際しての判断基準を示す（2.5 節)。

2.1　心拍活動

心臓の拍動する回数を心拍動（HR: Heart Rate）という。一般的には 1 分間の拍動数（BPM: Beats Per Minute、回/分）で計られるが、目的によっては RR 間隔時間（単位: 秒）も用いられる（$\langle \text{BPM} \rangle = 60/\langle \text{RR 間隔} \rangle$ で相互に換算)。安静時の成人の心拍数は 60 BPM – 70 BPM だが、年齢、性差、体位、運動、薬物、気温、精神活動によって変化する [37]。本研究で関心があるのは、精神活動による変化である。

心拍動は、心臓右心房上端に位置する洞房結節（sinoatrial node）が定期的に放電し、この放電とそれに伴う心臓筋肉の収縮が心臓全体に遅延を伴いながら上方から下方に伝播することで発生する。この放電リズムは自律神経（交感神経と副交感神経）及び中枢神経の活動によって変動する [5]。副交感神経（PNS: Parasympathetic Nervous System）の賦活は心拍動の低下を、交感神経（SNS: Sympatetic Nervous System）はその逆に賦活によって心拍動上昇をもたらす。心臓には恒常機能があり、例えば、脳内の血圧低下は首に位置する血圧受容器（baroreceptor）が検出し、この器官が SNS を介して心拍動上昇を促す。基本的には、心拍動の上昇はこのように SNS の賦活によってもたらされるが、SNS が一定でも PNS の活動が後退することによっても発生する。すなわち、心拍動は SNS と PNS のバランスによって変化する。

交感神経系の活動は、心拍変動（HRV: Heart Rate Variability）に反映するとされる [82]。心拍変動とは心拍動のゆらぎで、心拍動の時系列データ（RR 間隔時間が用いられる）を周波数領域に変換することで求められる。このとき、パワースペクトラムには図 2.2 のように二つのピークが現れる。0.04 Hz – 0.15 Hz 帯を低周波（LF: Low Frequency）成分、0.15 Hz – 0.40 Hz 帯を高周波（HF: High Frequency）成分という。LF 成分は血圧性の変動と呼ばれ、SNS と PNS 双方（主として SNS）の活動を反映するとされる。LF 成分はまた、心的負荷の指標としても知られている。後者は呼吸性の変動と呼ばれ、PNS の活動を反映するとされる。

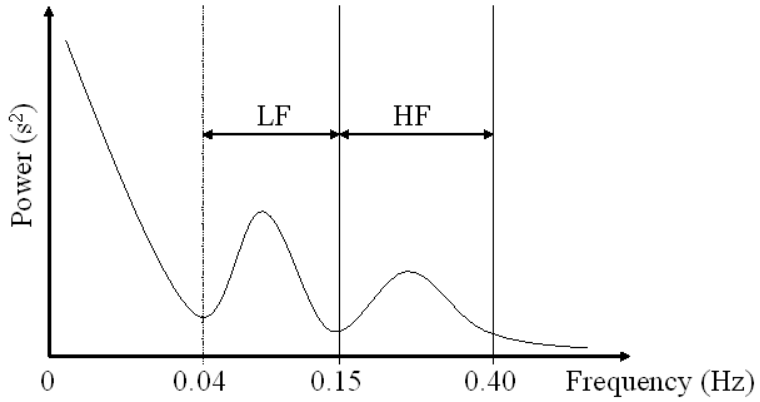

図 2.2: 心拍変動（HRV）の主要周波数帯

2.1.1 無拘束型心拍センサの動向

本研究では、目的とする要約映像の生成に望ましい性質を有しているという以外に、無拘束な測定が可能であるという理由から、心拍活動を指標に採用している。ここでは、こうした無拘束型の心拍測定技術を概説する。

HR、すなわち RR 間隔時間及び 1 分間の R 波出現回数（BPM）は、心臓の電気的興奮を体表面に装着した電極によって記録する心電計（ECG: Electro-Cardio-Graphy）で測定するのが特に医療目的では今も一般的である。しかし、人間工学やヒューマンインタフェース目的ではその拘束性が問題になることが多いため、無拘束・非侵襲型のセンシング技術が検討されてきた。

こうした心拍センサの多くは、血液を送り出す心臓の圧力の反作用で変化する体動を測定する心弾波計（BCG: Ballisto-Cardio-Graphy）の原理を利用している。例えば、新関らのベッド埋め込み型のセンサは胸郭位置に圧電素子 8 個を配置することで、心電図との比較で誤差 7.3% の精度を示している [62, 63]。同様に、Anttonen らは座面、背面、アームレストに感圧フィルムを貼り付けた椅子型センサを製作しており、光電脈波計との比較で $r = 0.99$ という精度を示している [6]。更に、鈴木らは椅子背後に設置したマイクロ波レーダーにより心拍動起因の微小な体動を計測することで、心電図と比較して $r = 0.96$ の精度を達成している [78]。

これらの技術を利用すれば、映像視聴時の心拍活動を視聴者に負担をかけずに実用的な精度で測定することが可能である。

なお、本研究は心拍動の無拘束技術を評価するものではないので、評価実験には ECG タイプの簡易測定装置（Polar S810i [69]）を用いた。この装置は二つの電極を持つバンドとそこから無線発信されるデータ信号を受信する腕時計型の装置で構成されている。測定精度は ± 1% 以内、測定レンジは 30 BPM – 240 BPM で、RR 間隔がミリ秒単位で受信機に記録される。記録データは Windows PC に付属のケーブルで転送でき、テキスト形式で読むことができる。実験では、バンドを被験者の胸部に直接装着し、受信機は無線到達範囲内（数メートル）に設置した。なお、装着後数秒間は、自動キャリブレーションのために測定はできない。

2.2　理解しやすさと心拍活動

1.2 節で示したように、本研究の目的の一つは理解しやすい映像区間を集めた要約映像の生成にある。ここで理解のしやすさは、 認知処理に関わる心的負荷に関係していると考える。例えば、注意を逸らせるような冗長な情報が多かったり、画面中の注視対象が分散していたり頻繁に切り替わったりする映像内容は、処理により多くの心的リソースを費やす。そして、このように複雑な画面は、一般には理解しづらいものと感じられる。そこで、ここでは理解しやすいとは、映像内容を認識するに際しての心的負荷が低い状態とする。

心的負荷は、心拍変動低周波成分（LF 成分）に反映するとされる [37]。心的負荷に対する LF 成分の挙動を模式的に図 2.3 に示す。LF 成分の低下は、例えば記憶文字と比較文字をマッチングさせる記憶タスクで記憶文字数を増加させる [1]、航空管制時に処理すべき機体数を増加させる [71] のように、心的負荷を加えたりワーキングメモリを消費するタスクを課すと生じると報告されている。また、長期間の心的負荷状態が続くと、恒常的に LF 成分が低下することも知られている [70]。但し、過度な心的負荷が加えられると、個体は例えば理解や記憶をあきらめるため、LF 成分は増加に転ずる。しかし、一般的な映像では理解をあきらめるほど理解しがたい映像区間はさほどないと考えることができるので、ここでは LF 成分は心的負荷に対して単調減少すると仮定する。

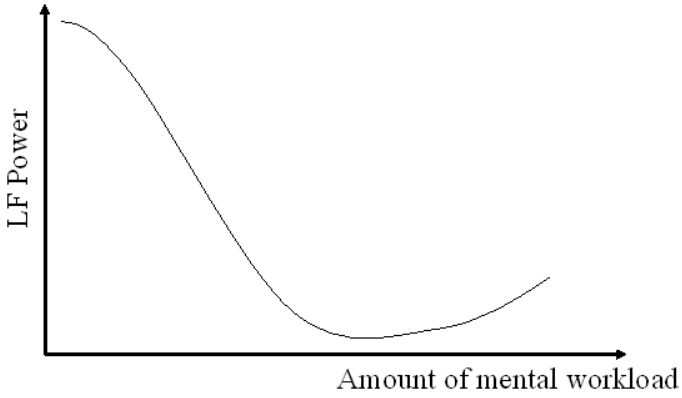

図 2.3: 心的負荷に対する LF 成分の挙動

LF 成分の変動は主に SNS の活動を反映したものだとされているので、心的負荷の高い状態を興奮状態、すなわち SNS の賦活と考えると、興奮で LF 成分が抑制されるのは背反しているように見える。これについては、Task Force of the European Society of Cardiology らは、心拍変動の全エネルギーの低下に伴って LF 成分も減少するため [82]、機械システム振興協会はその報告書で、心的負荷時の SNS の賦活と PNS の後退は平均レベルについて言えることであり、実際には無視できないレベルのゆらぎが存在するからと説明しているが [42]、心的負荷と LF 成分の関係は自律神経活動の観点からは十分に説明付けられていない (e.g. 杉田ら [77])。本論文ではこの関係性の説明は差し控え、心的負荷が LF 成分に逆相関している現象面に着目し、これを映像要約に応用することに専念するというアプローチをとる。

LF 成分を用いた理解しやすい映像区間の選択方法の概略を以下に示す。まず、時系列的にばらつきのある RR 間隔イベントデータは、アーティファクト除去後、スプライン関数を用いて等時系列に補間する。次に、この等時系列 RR データを一定幅のウィンドウ（$[t, t+L]$。但し t は映像・データ時刻、L はウィンドウ幅）に分割し、それぞれに窓関数処理を施す。窓関数には、生体信号のパワースペクトラム解析で一般的に用いられるハニングウィンドウを利用する [39]。続いて、窓関数処理後のウィンドウに離散フーリエ変換（DFT）を施し、0.04 Hz – 0.15 Hz 帯のパワー和を算出して LF 成分値とする。但し、刺激に対する LF 成分の反応は 1 秒 – 3 秒遅れるとされているので [71]、映像時間中の刺激発生時点と指標値を同期させるため、得られた指標値を時

間軸上でずらす。映像区間が RR データウィンドウと一致していれば、対応するウィンドウの LF 成分値を基に映像区間を選択すればよい。しかし、区間単位がショットの場合は、まずショットとは無関係に RR データウィンドウ単位で選択を行い、その上でウィンドウに属するショットを補助的な指標を用いて抜粋するという手順が必要になる。この補助指標には、心拍動と LF 成分との間に弱い相関があるという知見 [1] を基に、心拍動（BPM）データを用いる。具体的な信号処理方法は、第 3 章で説明する。

2.3　印象と心拍活動

本研究のもう一方の目的は、印象的な映像区間を集めた要約映像の生成にある。ここで印象的であるとは、辞書的には、興奮、注意、畏敬の念、感嘆をそれを観るヒトに与える対象物の持つ力の度合いを指す。つまり、印象的な映像区間とは、観ることで視聴者が興奮したり熱狂的になったり、感じ入ったり、緊張したり、眼の覚めるような思いを強く経験させる力を持つ区間のことである。こうした心的状態は、情動の動機付けモデルが定義するところの、覚醒度に対応すると考える。覚醒度の高まりは心拍動の低下と副交感神経系の賦活に現れるとされているので、これら二つの状態を検出できれば、覚醒的＝印象的な映像区間を検出できる。

2.3.1　印象と心拍動

覚醒度は、情動の動機付けモデル（motivational model of emotion）が説明する情動を構成している要素の一つである [16, 17]。このモデルは、情動及び情動を誘起する刺激を覚醒度（arousal）と情動価（valence）という 2 要素で構成している。情動価は刺激に誘起される行動のレパートリを選択する要素で，快 – 中性 – 不快（pleasant – neutral – unpleasant）又は接近 – 逃走の軸で示される。覚醒度は行動の強度を決定する要素で、高覚醒 – 低覚醒又は興奮 – 沈静（excited – calm）の軸で示される。例えば、強い覚醒度と強い不快さ（逃走）を持つ外界の刺激は、個体に急速な逃走を促す。（覚醒度, 情動価）の組で示される情動を日常的な語に置き換えると、表 2.1 のように

表 2.1: （覚醒度, 情動価）の組で示される情動に対応する日常的な形容詞（Lang, et al. [44] より）

Emotion space	Words
Valence - pleasant	happy, pleased, satisfied, contented, hopeful
Valence - unpleasant	unhappy, annoyed, unsatisfied, melancholic, despaired, bored
Arousal - excited	excited, frenzied, jittery, wide-awake, aroused
Arousal - calm	relaxed, calm, sluggish, dull, sleepy, unaroused

なる。

心拍動（HR）は、覚醒度の強度に応じて低下するとされる [5, 6, 16, 17, 66, 88]。そこで、覚醒的＝印象的な映像区間を検出するには、HR の低下を検出すればよい。

但し、Bradley らの静止画像を用いた実験によれば、HR の低下パターンは図 2.4 に示したように刺激画像の情動価に応じて異なる [16] 。HR は、不快な画像で下降と上昇という 2 相的な、快な画像及び中性な画像で下降 – 上昇 – 下降 – 上昇という 4 相的なパターンを描く。HR 低下の度合いは不快な画像が最も大きく、続いて快な画像、中性な画像の順になる。映像の情動価はあらかじめ知ることができないので、検出アルゴリズムには情動価に依存しない方法が求められる。

そこで、本研究では以下のような処理方法を HR データに適用することで、印象的な映像区間を選択する。まず、HR の時刻 t での指標値は、ウィンドウ $[t, t+6]$ の平均で代表させる。このウィンドウ幅は、刺激呈示時点（t）から 6 秒程度までの間で HR が基準値以下に低下しているのが三つの情動で共通していることから決定した。また、この 6 秒ウィンドウは、予備実験で始点と終点を何パターンかで検証することで、最も主観的な判断に近いことが確認されている。しかし、これだけでは、この固定長ウィンドウが可変長のショットと一致しない。そこで、この 6 秒ウィンドウを、必要な時間精度単位でずらすスライディングウィンドウにする。これは、移動平均処理であるため、得られるカーブの形状は比較的スムースになる。そして、このカーブの極小点を含むショットを覚醒的な映像区間とし、指標値を割り当てる。具体的な信号処理方法は、第 4 章で説明する。

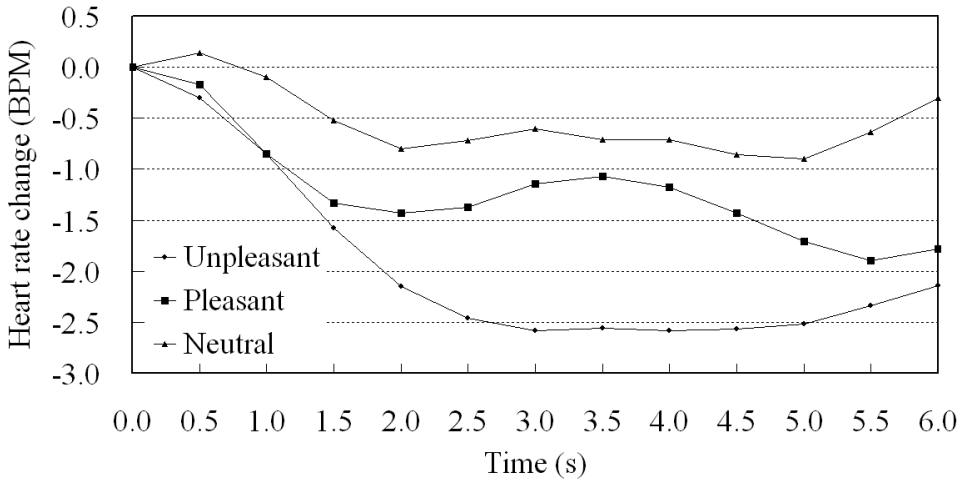

図 2.4: 異なる情動価（快／中性／不快）の静止画が呈示されたときの心拍動変化（Bradley, et al.[17] より）

2.3.2　印象と副交感神経活動

交感神経系の活動と外界の刺激が呈示されたときの生体の反応との関係は、防衛の段階的反応モデル（defence cascade model）で説明されている [16, 17, 66]。

このモデルは、生体の反応を 2 段階に分けて説明する。まず、外界刺激を感受した最初の段階では、生体はその処理に専念し、まだ行動には至らない（定位反応。OR: Orienting Reflex）。このとき、刺激処理に心的リソースが費やされるため、他の活動は抑制される。例えば、心拍動が低下し、皮膚電気反応が増加し、驚愕反応が抑制される。また、SNS と PNS 双方が賦活する。但し、心拍動が低下することからもわかるように、PNS の方が優勢になっていると考えられる。外界刺激は先に説明した（覚醒度, 情動価）空間にプロットされ、その情動価に応じて行動のレパートリ（刺激から逃避するか接近するか）が、覚醒度によってその強度（どれだけ早く行動するか）が決定される。

その後、生体は行動段階に移行する。刺激が存在に危機をもたらすような不快で覚醒度の高い事象ならば、個体はその事象から急速に逃走する。快で覚醒度の高い事象は生命維持をサポートする事象（食物や生殖対象）なので、急速な接近をもたらす。こうした実際の行動をサポートするため、この段階では SNS が賦活する。すなわち、心拍動を上昇に、驚愕反応を活性化に転じることで、急速な身体運動が可能なように準

備する。但し、映像視聴で行動段階に移行するのはまれであり、心拍上昇に転じるのは恐怖症のある視聴者がその対象となる映像を視聴したときぐらいであるとされる [16]。したがって、映像を対象とする本研究では、第 1 段階の生理心理学的挙動にのみ着目すればよい。

覚醒的な刺激と対峙したときに心拍動が低下するというのは、驚いたときに鼓動が早くなるという我々の日常の経験とは背反するように思われる。しかし、このことはこの 2 段階モデルで説明でき、鼓動の高まりは行動段階への移行を指している。実際、第 1 段階と第 2 段階は混同しやすく、Drescher らの実験では、背後から唐突に挨拶を受けた被験者は鼓動の高まりを報告したが、第 1 段階の外界刺激受容時には心拍動が低下したという結果が得られている [21]。

本研究ではこのモデルを援用することで、映像区間によってもたらされる覚醒度＝映像区間の印象度を PNS の賦活、すなわち心拍変動高周波成分（HF 成分）の増加から検出する。HF 成分を用いた印象的な映像区間の選択方法の概略を以下に示す。

まず、LF 成分値を算出した方法（2.2 節）同様に、固定ウィンドウ幅（L）の心拍動にハニングウィンドウをかけた上で DFT を施し、HF 帯である 0.15 Hz – 0.40 Hz のパワーの和を求める。このとき、時刻 t の HF 成分値の算出元となるウィンドウの区間を $[t - L/2, t + L/2]$ とする。そして、このウィンドウを必要な時間精度単位でずらすスライディングウィンドウにする。LF 成分の処理方法と異なるが、これは、どちらも印象度の指標である本 HF 成分指標と先の HR 値指標を最終的には組合せて利用するため、手法を揃えたかったためである。

ここで、LF 成分のような刺激に対する反応の遅れや、HR のような情動価別の時系列的な挙動を考慮する必要があるが、HF 成分若しくは PNS の活動については、遅延や情動価との関係は知られていない。そこで、HF 成分の増加は図 2.4 で HR カーブが上昇に転じる 4 秒程度まで持続していると仮定し、時刻 t の指標値は、先に得た HF 成分値の $[t, t + 4]$ の間の平均値とする。そして、この移動平均処理によって得られたスムースなカーブの極大値を含むショットを覚醒的な映像区間とする。具体的な処理方法、及び HF 成分から得た指標を HR の指標値と組合せる方法は、第 4 章で説明する。

2.4　実験用刺激映像の検討

映像要約技術は映像のジャンルやタイプに依存することが多く、あるタイプで有効な方法も他のタイプでは不適切であることはしばしば起こる。しかし、既存の経験型映像要約研究には、提案方法がどのようなタイプの映像に適用可能かが明確に示されていないものが多い。当然ながら、提案方法がどのようなタイプになら適用可能なのかを明らかにするには、評価実験で使用する刺激映像のタイプをあらかじめ分類しておく必要がある。本節ではこの分類基準を説明した上で、第 3 章以降で使用する刺激映像を示す。

本論文では、映像のタイプは 2.3.1 節で説明した情動の動機付けモデルに即して分類する。評価実験では、例えば、HR を用いた印象的な映像要約方法は（覚醒度＝高, 情動価＝快）なタイプの映像で効果的だが、（覚醒度＝低, 情動価＝中性）なタイプでは検出精度が低い、のように示す。

これには、覚醒度と情動価が異なる数種類のオリジナル映像が実験刺激に必要になるが、（覚醒度, 情動価）で構成される情動空間（emotion space）の代表的な点を占める刺激を選択すれば、効率的に実験を行うことができる。各種の静止画像を用いた Lang らの実験によれば、情動空間で刺激画像は一様に分布せず、覚醒的で快な点から非覚醒的で情動価が中性な点を結ぶライン上と、覚醒的で不快な点から非覚醒的で情動価

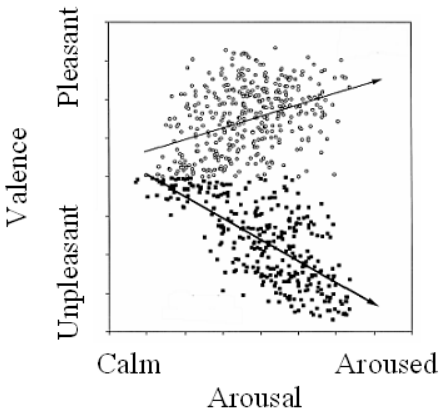

図 2.5: （覚醒度, 情動価）空間における静止画像刺激の分布 – 図中の直線は快／不快でデータを分けたときのそれぞれの回帰直線を示す（Bradley, et al.[17] より）

表 2.2: 実験用刺激オリジナル映像

Type	Length (s)	Arousal	Valence	Contents
A	412	Low	Neutral	A bird documentary film
B	418	High	Pleasant	A football game with three goal scenes
C	377	High	Unpleasant	A short surealistic film

が中性な点を結ぶライン上に、「く」の字状に分布する [17, 44]。この分布の様子を図 2.5 に示す。そこで、この空間全体の特徴は「く」の字の回帰直線の三つの節で代表させることができるとし、実験刺激映像には（低覚醒, 中性）、（高覚醒, 快）、（高覚醒, 不快）という特徴を持つ映像を選択する。以下、これらを順に映像 A、B、C と呼ぶ。

3 本の刺激オリジナル映像は、Bradley らの実験 [17] を参考に実験者が選んだ（表 2.2）。いずれも、元資料映像から実験のために連続した 7 分程度を切り出した、オーディオを伴うカラー映像である。平均ショット数は 60 である。映像 A は野鳥のドキュメンタリー映画で、野鳥の飛翔を前半では BGM を、後半では背景音を伴って淡々と追った映像である。映像 B はサッカーゲームで、中盤にゴールシーンのリプレイ（スローモーション）、最後に新規のゴールシーンとそのリプレイ（スローモーション）といった印象的なイベントを含むように切り出したものである。映像 C は眼球をボタンかのように指で押したり口に手を突っ込むなど、不快感は高いが刺激的な内容を含む短編映画である。いずれの映像でも言語情報を含まないようにしたが、映像 B ではスコア表示や選手交代時の氏名表示が映像に含まれている（別トラックに収録されている解説音声は省いた）。

上記刺激オリジナル映像の情動空間中の位置を確認する目的で、画像の覚醒度値と情動価値を 5 件法で主観評価した。映像を呈示した被験者への質問には Lang らの International Affective Picture System（IAPS）の質問票 [44] を用いて、情動価が快は「楽しい、喜ばしい、満足した、希望的」、情動価が不快は「楽しくない、気に障る、いらいらする、憂鬱な、飽きた」、高覚醒は「興奮した、熱狂的な、緊張した、めまぐるしい、目の覚めるような」、低覚醒は「落ち着いた、静かな、活気のない、鈍い、眠気

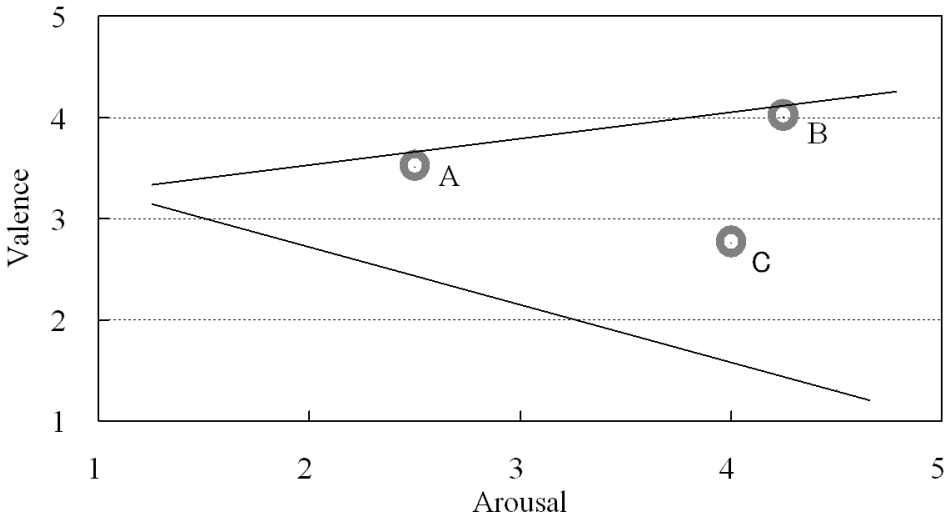

図 2.6: 実験用オリジナル映像の情動空間中の位置（図中の直線は図 2.5 に示した Bradley, et al. [17] の情動空間における刺激分布の直線近似）

を催す」といった経験・感覚であると説明した[1]。結果を 図 2.6 に示す（図中の直線は図 2.5 の回帰直線)。映像 A と映像 B はおおむね実験者の意図通りにそれぞれ（低覚醒, 中性)、(高覚醒, 快）に位置したが、映像 C は高覚醒ではあるが情動価が中性寄りになっている。これは、映像 C が不快と想定される映像だけで構成されていないためと考えられる（実験者のカウントでは 20% 程度)。しかし、不快さを意図した映像であっても、不快な場面だけで全編が構成されている映像は非常に特殊なものと考えられるため、一般的な視聴対象としてはこの程度の位置が、くの字下端の限界であると考える。

2.5　評価基準

映像要約の評価方法には、要約映像の内容や目的に応じて、1）提案方法が選択した区間と視聴者が主観的に選択した区間との比較、2）提案方法が選択した区間と専門家若しくは映像作成者本人が主観的に選択した区間との比較（e.g. 講義映像では講義者本人、テレビ番組では制作プロデューサ、スポーツ番組ではニュース番組のハイライト)、3）被験者を用いた、選択区間が適切であるかの主観評価実験、4）要約映像視聴

[1]原文は英文（表 2.1 参照）であり、訳は実験者が当てた。

後にタスクを課し、タスク達成度で評価する実験（e.g. 講義映像や教育番組）などがある [85]。本研究では、1.2 節に示した目的の性質を考慮し、上記 1 と 3 の方法と映像構造を基にした方法で提案方法を評価する。

評価基準 1（適合率） 提案方法で生成した要約映像を、その要約映像の生成に利用した心拍活動データを提供した被験者が主観的に選択した映像区間を集めた要約映像を比較する（上記 1）

評価基準 2（主観評価） 提案方法で生成した要約映像を、対照となる要約映像と共に被験者に呈示することで主観的に評価する（上記 3）

評価基準 3（構造評価） 提案方法で生成した要約映像とオリジナル映像からそれぞれ映像内から抽出できる指標を比較する

理解しやすい要約映像（第 3 章）では評価基準 2（主観評価）と評価基準 3（構造評価）を基に、印象的な要約映像（第 4 章）では評価基準 1（適合率）と評価基準 2（主観評価）を基にそれぞれ評価する。

適合率は、Smeulders らの定式に従い、被験者へのオリジナル映像の呈示を通じて指定の判断基準で主観的に選択させた映像区間（R）と本提案方法が選択した区間（A）から以下の式で求める [74]。

$$p = \frac{|A \cap R|}{|A|}$$

つまり、本提案方法が選択した区間で主観的選択の中にもある区間の数を、全体の区間数で除したものが適合率 p である。なお、本論文では提案方法と主観的選択とで抽出する区間数を一致させている（10 ショット）。そのため、適合率（precision: 上式）と再現率（recall: 上式の分母が R）は同じものになる。

得られた適合率がどの程度であれば方法が適用可能であるかの判断は、表 2.3 に示した生理心理学的指標を用いた既存の情動分類手段の適合率、特に高橋 [79] の研究を参考にする（表中のランダム適合率は、例えば 25% とは、被験者の情動状態を 4 種類のカテゴリーに分類したときに、無作為な選択時の適合確率を示す）。生理心理学的手法

表 2.3: 生理心理学的指標を用いた情動分類手段の適合率

Method	Hit rate (%)	Random hit (%)
β wave †	43.3	Not shown
Face muscle *	38 – 51	25.0
Brain wave ‡	35.0 ± 19.0	20.0
Blood pulse ‡	26.7 ± 16.1	20.0
GSR ‡	15.0 ± 12.2	20.0
Brain wave + blood pulse ‡	28.3 ± 12.2	20.0
Blood pulse + GSR ‡	16.7 ± 5.3	20.0
GSR + brain wave ‡	31.7 ± 6.2	20.0

†Miyata 2004 [53]
* Picard 2001 [68]
‡Takahashi 2005 [79]

に基づいた既存の映像要約研究（1.3.4 節）を参考にしなかったのは、適合率が明示的でないものが多かったためである。表 2.3 から、平均して 30% の適合率が得られれば、本提案方法は他の生理心理学的指標を用いた方法と同程度の検出能力があると判断する。より高い適合率が望ましいのは当然だが、本方法は無拘束に生理心理学的指標の取得が可能という他の方法にはないメリットを有するため、同等以上ならば実用的には有利であると考える。

当然ながら、これは、30% 程度の適合率で要約映像が実用に耐えうるレベルであるということ述べるものではない。他の映像要約方法と同様に、生理心理学的指標を用いた映像要約も、複数の方法と補完しあうことで精度を高めることができると考える。このときの方法は生理心理学的なものでなくとも、映像内情報を利用したもの（1.3.3 節）であってもよい。しかし、方法の組合せに際しては、それぞれの方法の適用範囲とメリット・デメリットが明らかになっている必要がある。本研究はこうしたときの基礎的な知見を提供することで、映像要約研究に貢献する。

2.6 まとめ

本章では、本研究の前提となる心拍活動の生理心理学的知見を説明し、その応用方法を示した。また、評価実験で使用する刺激映像を情動分類に従って選択し、評価基準を述べた。本章をまとめると次のようになる。

- 目的の一つである「理解しやすい」要約映像については、理解しやすいということを視聴時の心的負荷が少ないと解釈し、心的負荷と関係があるとされる心拍変動低周波成分（LF 成分）を指標に、区間選択を行う。
- もう一方の目的である「印象的な」要約映像については、印象が情動の動機付けモデルでいうところの覚醒度に相当するとし、覚醒度と関係があるとされる心拍動（HR）と心拍変動高周波成分（HF 成分）を指標に、区間選択を行う。この 2 指標は、後に組合せて利用する。
- 本研究が心拍活動に着目したのは目的に適した特性を備えているだけでなく、無拘束な測定手段が実用レベルに達しているからである。現在の技術の測定精度は $r = 0.96$ 以上又は誤差数 % であり、十分に実用的であることが示されている。
- 本研究の提案方法の適用範囲を明確にするため、評価実験で用いる刺激には、異なる（覚醒度, 情動価）の値を持つ映像 A（低覚醒, 中性)、映像 B（高覚醒, 快)、映像 C（高覚醒, 不快）の 3 種類の映像を用いる。また、生理心理学的指標を用いた情動分類手段を参考に、本方法が 30 % 以上の適合率を示せば、少なくとも他の方式と同等の精度があり、無拘束性というメリットを考えると、実用上は優位であると判断する。但し、これは本方法単体で実用的な映像要約が可能であることを示すものではなく、他方法と組合せるときの基礎的知見として応用されることを期待するものである。

第3章　理解しやすい要約映像

3.1　目的

本章では、1.2 節で目的に掲げた「理解しやすい」要約映像の生成方法を検討する。ここでは要約映像を、以前観た映像を思い起こすときのメモや日記代わりに用いたり、その映像について知らせたい第三者にメールなどで送付することでどのような映像を観たのかを簡単に相手に示すメッセージとして利用することを想定している。以下、こうした想定用途を前者については「映像メモ」、後者については「映像メッセージ」と呼ぶ。こうした用途では、受け手の時間を長く束縛することのない、またシステムやネットワークに負担をかけないサイズが求められるため、非常に短い要約映像が求められる。しかし、極度に短くすると、その内容がわかりづらくなる。そこで、オリジナル映像から映像区間を選択するに際しては、理解のしやすさを基準とする必要がある。

上記の要求を満たすため、区間選択には映像外情報／心理生理学的反応を利用する（1.3.1 節）。具体的には、2.2 節で説明したように、映像区間が理解しやすいということは映像視聴に伴う心的負荷が少ないと仮定し、心的負荷の指標として知られている心拍変動低周波成分（LF 成分）を利用する。

但し、理解しやすい場面が内容的に代表的であるとは限らないため、理解しやすいという観点だけで要約映像を生成するのは好ましくないかもしれない。そのため本提案方法は、既存の方法（1.3 節）で特定の観点から粗く区間選択を行った後で、選択区間の中でも理解しやすい区間をピックアップする 2 次処理として利用するのが実用的と考える。但し、本章は LF 成分を用いて選択した区間が他の区間よりも理解しやすいことを検証するためのものなため、評価に影響を与える恐れのある恣意的な 1 次的な処理は行わない。

評価実験では、2.4 節で示した映像 A（低覚醒, 中性）、映像 B（高覚醒, 快）、映像 C（高覚醒, 不快）を呈示刺激に用い、2.5 節で示した評価基準 2（主観評価）と評価基準 3（構造評価） を基に評価する。評価の対照には、主観評価ではオリジナル映像から等間隔に選択した区間を用いた要約映像を、構造評価ではオリジナル映像を用いる。

以下、3.2 節で映像要約方法を説明した上で、3.3 節で要約映像の生成とその評価実験について、3.4 節でその結果を示す。最後に、3.5 節で得られた知見をまとめて示す。

3.2　映像要約方法

心拍変動 LF 成分を利用した映像要約方法を、1.3.2 節で示した一般手順に沿って示す。オリジナル映像はまず固定区間長に分割し、そこから区間選択手順で理解しやすい区間を選択する。続く区間短縮手順では、区間に属する複数のショットのうち 1 ショットを抜粋する。続く連結手順では、連結上の処理や効果を含めると生成した要約映像の理解度に影響を与える恐れがあるため、評価実験では特に処理は行わずにオリジナルの時系列順に連結する。

区間分割

オリジナル映像を、固定区間長 L で映像区間に分割する。以下、各区間を V_v（$v = [0, \mathcal{V}-1]$: 但し、$\mathcal{V}$ は全区間数）と記す。

区間選択

2.2 節で概説したように、区間 V_v にそれぞれ指標値を割り当てることで区間選択を行う。区間選択の手順を図 3.1 に示す。

最初に、映像視聴と同期して取得した未処理の RR 間隔データに対し、アーティファクト処理と等時系列化を施す（前処理）。アーティファクト処理では、RR 値が平均の $\pm 50\%$ を超えるデータを除去する。等時系列化では、リサンプリング周波数を f としたスプライン関数でデータ補間を行う。このとき、時間に対応する値には RR 間隔時

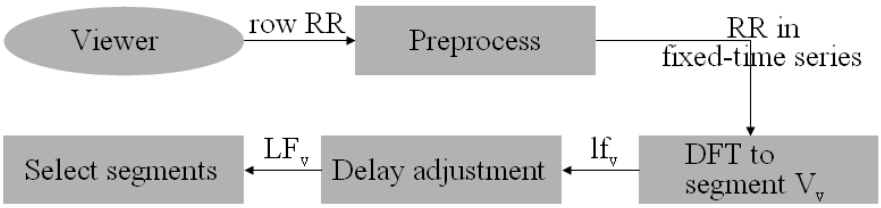

図 3.1: LF 成分を用いた区間選択手順

間を用いる。そして、この RR 等時系列データを映像区間と同じく L で等区間に分割する。

次に、V_v に属する等時系列 RR データにハニングウィンドウをそれぞれ施した上で離散フーリエ変換（DFT）をかけ、0.04 Hz – 0.15 Hz 範囲内のパワーの和を取ることで、区間単位の LF 成分値を得る。更に、この値の範囲が $[0, 1]$ となるように正規化する。この値を、区間 V_v の生 LF 値 lf_v とする。

続いて、刺激に対する LF 反応の 1 秒 – 3 秒の遅れ [71] を考慮し、lf_v を図 3.2 の要領で補正することで、区間 V_v の指標値 LF_v を得る。まず、指標値は基本的には lf_v であるが、区間始端時点（時刻なら $t = Lv$ 点）で呈示された刺激に対する反応は最短で 1 秒後に発生するので、lf_v の最初の 1 秒分（$1/L$）はこの区間に属するものではないとして差し引く。更に、区間 V_v の終端（$t = L(v + 1)$ 点）の刺激への反応が最長 3 秒

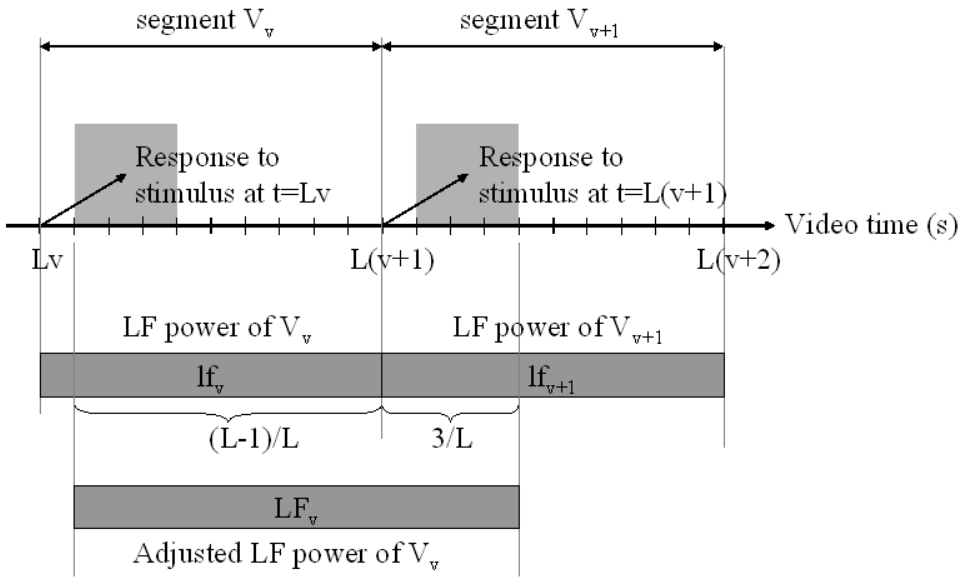

図 3.2: 映像区間に対応する LF 成分値を反応遅延を考慮して補正する方法

まで遅れるとして、隣接する区間 V_{v+1} の3秒分（$3/L$）の値を区間を加える。こうして得た補正 LF 値を、区間 V_v の指標 LF 値 LF_v とする。すなわち、

$$LF_v = \frac{L-1}{L} lf_v + \frac{3}{L} lf_{v+1}$$

本論文では、この LF_v 値が高い順に10区間を選択する。

区間選択 – 視聴者毎指標の集計

上記手順は視聴者1名の心拍活動データを基にしたものだが、複数の視聴者の指標値をまとめることもできる。これには、個々の視聴者（視聴者 $a = [0, \mathcal{A}-1]$。但し、$\mathcal{A}$ は視聴者数）の区間単位の生 LF 値を $lf_{v,a}$ とすると、次のように平均化したものを区間単位の生 LF 値とする。

$$lf_v = \frac{1}{\mathcal{A}} \sum_{a=0}^{\mathcal{A}-1} lf_{v,a}$$

後の遅延補正処理は個人単位と同じである。

この処理は、例えばグループで視聴したときに全体の指標値をまとめるときに用いる。

区間短縮

区間分割手順でオリジナル映像を固定長で分割したため、区間端には映像としては不自然な短いショットの断片が含まれることがある。そこで、選択された区間内からショットを一つ選択する。これには、LF 成分と HR の間に弱い相関があるという知見 [1] を利用し、HR の平均値を補助的に用いる。

区間からショットを抜粋する手順は次の通りである（図3.3）。まず、区間 V_v 内のカット点を、既存のカット点検出技術（1.3.2節）を用いて検出する。次に、HR の反応の遅延が LF 成分の1秒 – 3秒の中間である2秒だとして HR データをずらす。続いて、区間内のショット毎に平均 HR 値を求める（単位: BPM）。但し、ショットが区間端をまたいでいるときは、区間外のデータは用いない。最後に、この値が最も高いショットを選択する。このショットが区間端をまたいでいるときは、ショット全体を選

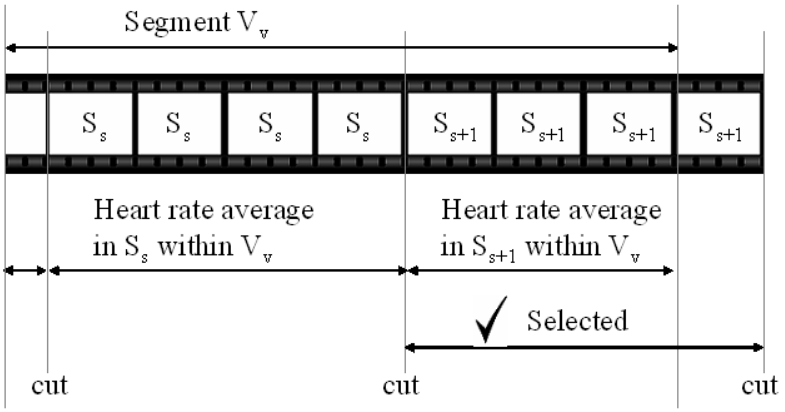

図 3.3: HR を補助的に用いて、選択区間 V_v からショットを選択する方法

択する。視聴者別の指標値を集計して区間選択を行う場合は、視聴者毎の HR の標準得点を求めた上で区間内ショットの平均を求める。

3.3 評価実験手順

以上の手順で 3 タイプのオリジナル映像から要約映像をそれぞれ生成し、主観評価実験と構造評価実験を通じて、本方法を評価した。

3.3.1 要約映像の生成

2.4 節で説明した 3 タイプ（A、B、C: 表 2.2）のオリジナル映像を被験者に呈示し、全被験者の HR データから前節の方法で要約映像を生成した。以下、この方法で作成した要約映像を「LF 版」とする。

図 3.4 に本実験のプロトコルを示す。

実験に先立ち、ECG 式の心拍計（2.1.1 節）を被験者の胸部に装着し、本実験の目的を簡略に説明した。但し、呈示する映像の時間長以外は、被験者にタイトルなど、映像内容については一切説明していない。そして、オリジナル映像の開始から終了まで RR 間隔データを取得した。

3 タイプの映像はそれぞれ画面に何も表示されない 2 分の安静期間を挟んで連続的に呈示したが、いつ映像が始まるか被験者が不安にならないよう、安静期間中の開始 60

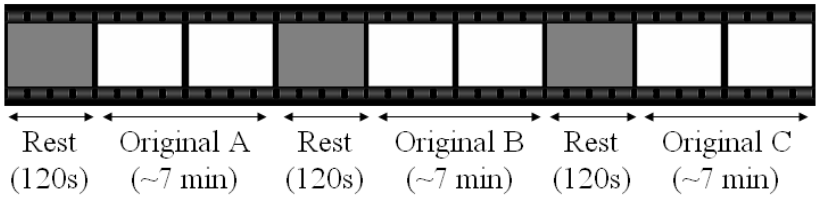

図 3.4: LF 版生成実験プロトコル – オリジナル映像 A、B、C の呈示順序は被験者毎にランダムに入れ替えた

秒点と 110 秒点で開始までの残り時間を 5 秒間呈示した。呈示順序は被験者毎にランダムに入れ替えた。呈示は、関心が他に逸れないように何もないパネルを背後に置いたカラーディスプレイで行い、被験者とディスプレイの距離が水平画角で 30° 程度になるように配置した。音量はいずれのタイプの映像でも同程度となるように、実験者が調節した。実験室内は比較的暗めの照明にし、室温は快適な程度に調節した。また、被験者には、姿勢が心拍活動に影響を与えないよう、リラックスした姿勢を取るように指示したが、同時に視聴中には極力体を動かさないようにも指示した。呈示中の咳や身体を掻くといったアーティファクトの要因となる体動は、実験者が記録した。

生理心理学的反応を用いた既存の要約映像作成実験（表 1.2）の被験者数が 2 名 – 10 名であることに依拠し、被験者数は 4 名とした。被験者は生理測定実験に慣れた、心臓疾患も喫煙習慣もない健康な 24 歳 – 32 歳（平均 26.5 歳）の男性である。実験後のインタビューによれば、被験者はいずれも映像 A と映像 C を視聴した経験はなかった。映像 B については、そのゲームを観たことは覚えているが、内容については記憶になかったと述べた被験者が数名いた（映像が公開されたのは実験の 4 年前）。

実験後、3.2 節の方法で要約映像を作成した。映像要約の区間選択手順では、記録された体動時点に突出した心拍動があれば、アーティファクト除去手順の前に手動で除外した。スプライン関数適用後のリサンプリング周波数は 2 Hz とした。区間長 L は、極力区間長とショット長を一致するよう、映画の平均的なショット長を参考に 10 秒を用いた[1]。また、選択する区間は 10 区間とした。区間短縮では、選択区間から各 1 ショットを選択した。なお、カット点検出は実験者が目視で行った。

[1]平均ショット長は 1960 年代までは 10 秒/ショット、1960 年代以降は 6 秒/ショット程度である [13, 14]。ここでは長めの時間を採用した。

3.3.2 主観評価

LF 版の理解しやすさは、被験者に LF 版と比較対照版を呈示し主観評価を行わせることで評価した。評価項目は次の 3 項目である。

1. 要約映像のまとまりはよいか（以下、「まとまり」）

2. 要約映像からオリジナル映像全体の話を推測・把握できるか（「ストーリー性」）

3. 要約映像を観たことでオリジナル映像を観たくなったか（「訴求性」）

項目 1 と項目 2 は映像の理解度に関わる設問で、項目 3 は映像メモ及び映像メッセージとしての本方法の効果を調査する設問である。メッセージを送るのは、それが映像である場合は特に、その映像を受け手にも視聴してほしいという意味が込められていると想定されるため、受信映像メッセージを視聴してオリジナルを観たくなるかは重要な特性と考えたからである。

主観評価実験で使用する要約映像のショット長は、心的負荷や理解しやすさに影響を与えないよう LF 版と比較対照版とで揃える必要がある。そこで、先に選択した各ショットの中から、HR 標準得点の平均がショット内でピークとなる時刻を中心に前後 1.5 秒を更に抜粋することで、全ショットが同じ長さとなるようにした（但し、3.2 節で示した区間短縮手順同様、2 秒の遅延を考慮する）。3 秒抜粋がカット点をまたぐ場合は、抜粋区間の端点をそこまでずらした（図 3.5）。なお、この 3 秒抜粋長は、映画予告編の TV スポットの平均的な時間長を参考に定めた。

比較対照版には、オリジナル映像を 10 区間に均等分割し、区間先頭からそれぞれ 3 秒ずつ切り出した抜粋を連結して生成した 30 秒の要約映像を用いた。但し、LF 版同様、抜粋区間内にカット点が含まれないよう、必要に応じて抜粋を前後にずらした。以下、これを「EQ 版」と呼ぶ。LF 版の理解しやすさの評点が EQ 版よりも有意に高ければ、本方法が妥当であると判断する。

被験者数は 19 名で、このうちオリジナル映像を視聴した経験のある被験者は 9 名（以下、既視聴グループ）、未視聴の被験者は 10 名である（未視聴グループ）。実験前

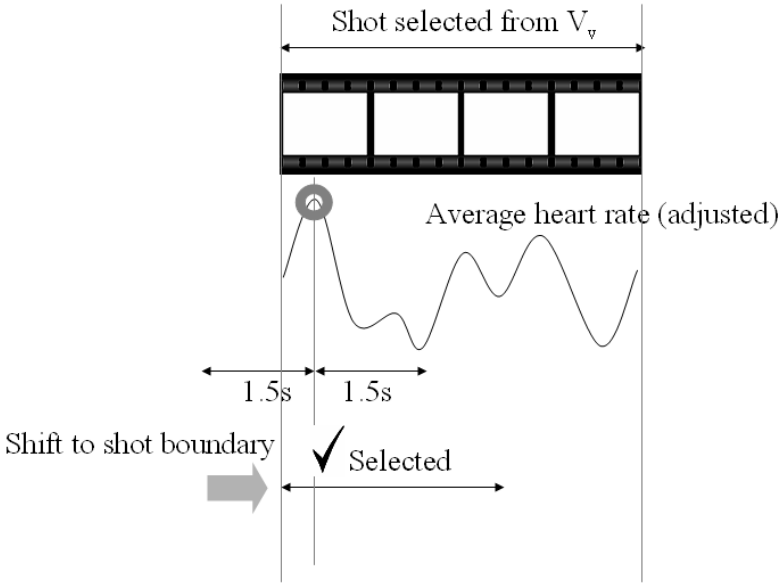

図 3.5: LF 版評価実験で、選択ショットから 3 秒の抜粋を抽出する方法

に、既視聴グループには以前視聴したことのある内容の要約版であることを、未視聴グループには内容を説明せずに要約映像である点だけを説明した。

図 3.6 に本実験のプロトコルを示す。各タイプの映像（図中、Type A、B、C の呈示順序は被験者毎にランダムに入れ替えた）の呈示は、それぞれ画面上に何も表示しない 90 秒の安静時間（図中、Rest)、2 種類の映像要約方法（図中 A1 と A2 で、ランダムに順序を入れ替えた LF 版と EQ 版を指す）の呈示、そして評価項目を画面に表示する 120 秒の回答期間（図中、Q）で構成した。評価項目への回答は 5 件法（i.e. 「まとまり」の場合、5=まとまりが非常によい、1=まとまりが非常に悪い）で口頭で行わせ、実験者がこれを記録した。映像の既視聴グループにはオリジナル映像の内容から判断して、未視聴被験者には印象だけで判断するように指示した。呈示環境は要約映像生成実験（3.3.1 節）と同じである。

3.3.3　映像構造の評価

LF 版の理解しやすさを、ショット長とロングショットの使用率という二つの映像構造で評価した。二つの評価基準は、それらが映像内容理解を促す演出上の技法として一般的に認められていることと、ワイプやズームインといった他の技法と比べて利用

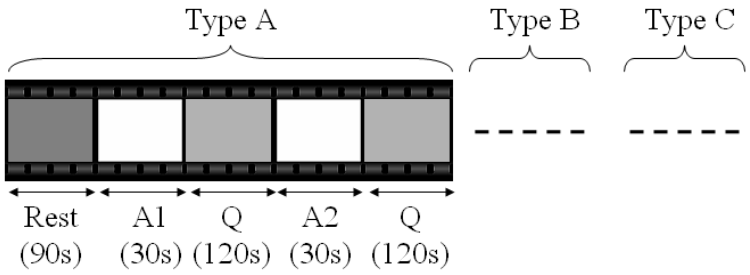

図 3.6: LF 版主観評価実験プロトコル – 映像タイプ（A、B、C）の呈示順序は被験者毎にランダムに入れ替えた。A1、A2 は LF 版又は EQ 版の要約映像で映像タイプ毎に順序を入れ替えた。

頻度が高いことから選択した。ショット長については、短ければ呈示内容を短時間で把握しなければならなくなるため [58]、長い方が理解しやすいと考えられる。ロングショットは、全体や続くシーンを映像的に説明するときに用いられるカメラワークであり、ロングショットの使用率の高い方が説明的な度合いが強く、内容把握が容易と考えることができる。

ショット長については、LF 版（但し、3.3.2 節で 3 秒長抜粋を行う前）の平均ショット長を求めた上で、これをオリジナル映像全体の平均ショット長で除すことでショット長の比（以下、ショット比）として求めた。ロングショットの使用率については、LF 版に登場するロングショットの数を全ショット数（10）で除すことでロングショットの登場比率を求め、これをオリジナル全体のロングショット登場比率で除して求めた（以下、ロングショット比）。なお、ロング、バスト、クローズなどのショットサイズは実験者が視認で行った。

3.4 結果と考察

主観評価実験と構造評価実験の結果を以下に示す。

3.4.1　主観評価

全評価項目／全映像タイプで主観評価の平均評点が、既視聴者グループと未視聴者グループで同じ傾向を示したことから、全視聴者グループについて評価項目別に平均評点を調べた。まとまりとストーリー性の結果を図 3.7 と図 3.8 にそれぞれ示す。LF 版の方が EQ 版よりも評点が高い傾向を示したのは、映像タイプ B だけである（$p < .05$）。このことから、主観評価の結果からは、本方法の適用範囲は（高覚醒, 快）な映像であり、他のタイプでは適用性が低いことがわかった。

この結果は、映像 A や映像 C のように言語情報を伴わない物語映像では、ショットのつながり（モンタージュ）によって形成されていたオリジナル映像の物語叙述が失われてしまったためと考える。つまりこれは、イベント志向だけで区間選択を行うと理解が困難になるということを示している。そうした場合、選択区間の前後数ショットを抽出するなど、連結手順上の方法を加味するとよい結果が得られると思われ、有効な連結手段を組み入れた検証が必要になる。また、映像 C については（無言劇で多分に暗喩的であるため）オリジナル映像の段階から比較的難解であったことも影響していると考えられる。このことから、元々難解な映像については本方法は適用しにくい可能性がある。

本方法が生成する要約映像の使用目的は映像メモ及び映像メッセージであるので、次に、こうした用途に適しているかを確認する。まず、評価項目毎／視聴者グループ毎に、全映像タイプについて集計した主観評価の項目を図 3.9 に示す。どちらのグループでも LF > EQ で LF 版の方が理解しやすいという傾向を示してはいるが、有意差若しくは有意傾向があるのは既視聴グループ（まとまりと訴求性で $p < .05$、ストーリー性で $p < .1$）である。このことから、既視聴の場合、すなわち一度自分で視聴したオリジナル映像の要約映像を個人的なメモとする用途では本方法は適切であるといえる。しかし、未視聴の場合、すなわち未見の映像の要約映像を第三者にメッセージとして送信したとき、受信側には必ずしも理解しやすいとはいえない、すなわち映像メッセージとしては有効性が限られることが示唆された。

映像メモとしてはどのタイプの映像に適切なのかについては、映像タイプ別の訴求

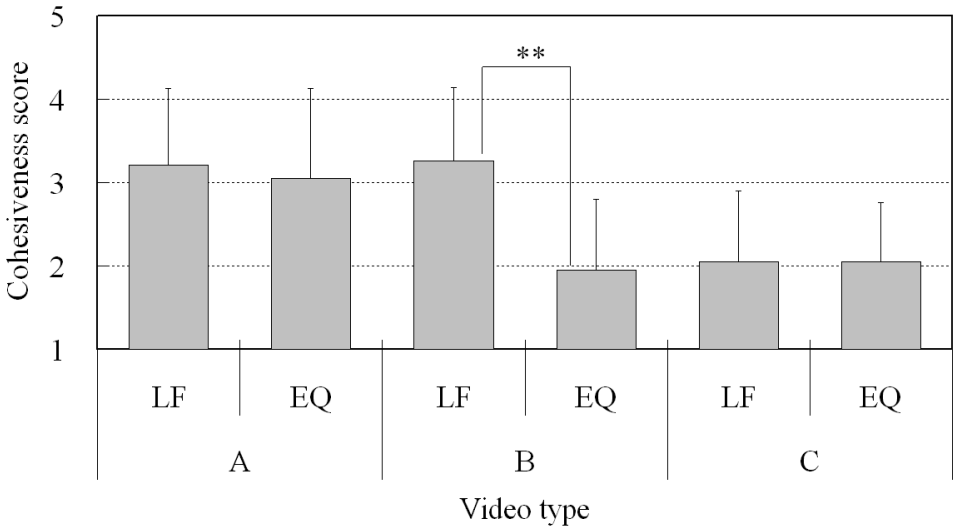

図 3.7: LF 版のまとまりのよさ（** $p < .05$）

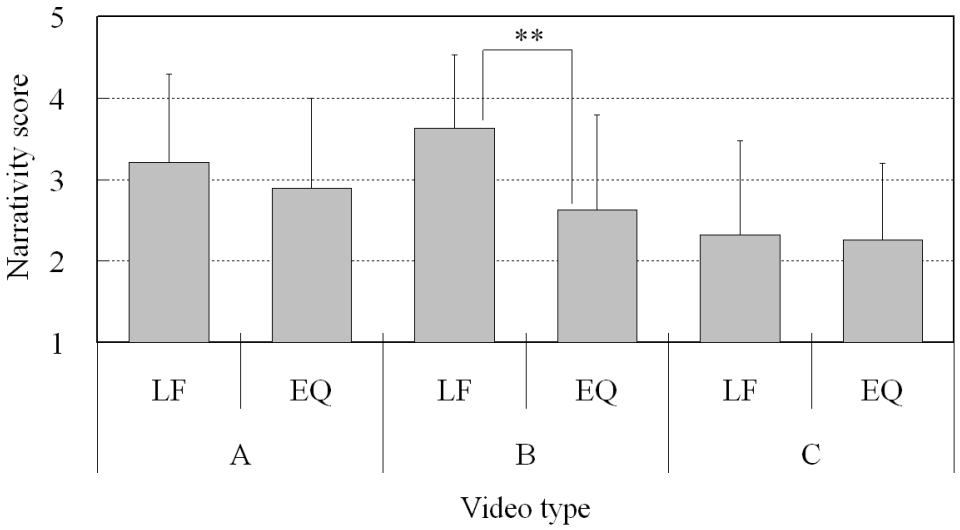

図 3.8: LF 版のストーリー性（** $p < .05$）

性評点から調べた。既視聴者の映像タイプ別の評点平均を図 3.10 に示す。この結果から、訴求性を有する映像タイプは（低覚醒, 中性）の映像 A（$p < .05$）と（高覚醒, 快）の映像 B （$p < .1$）である。

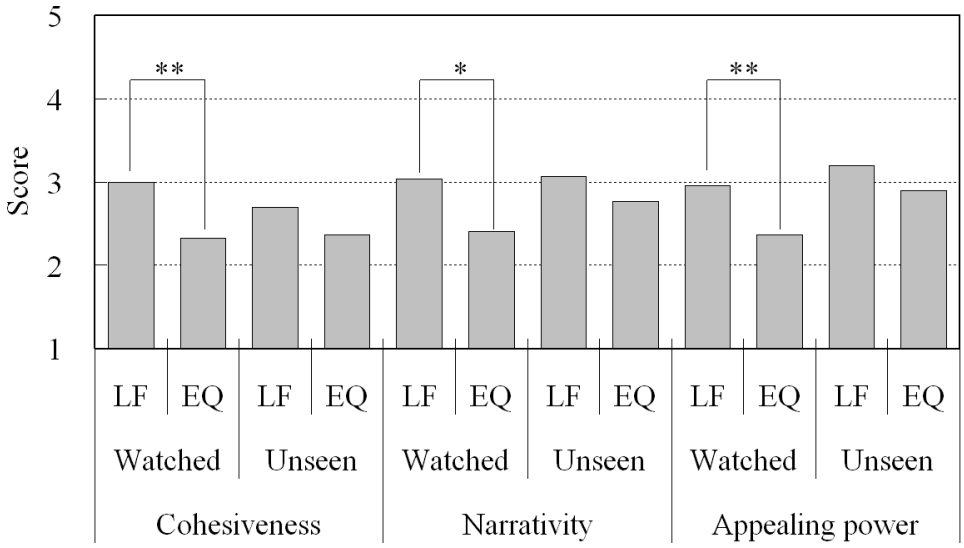

図 3.9: LF 版における被験者グループ別主観評価平均評点（** $p < .05$, * $p < .1$）

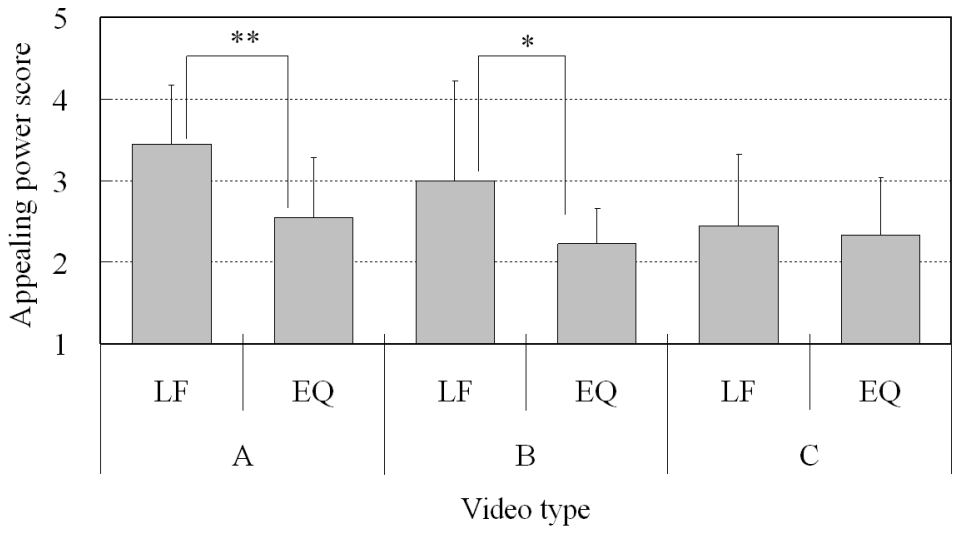

図 3.10: LF 版の既視聴者への訴求性（** $p < .05$, * $p < .1$）

3.4.2　映像構造の評価

ショット比とロングショット比の映像タイプ毎の平均を、図 3.11 に示す。ショット比では、映像 A と映像 C で LF 版の方がオリジナル版より長いことが示されている（それぞれ $p < .05$、$p < .1$）。ロングショット比では、映像 B で LF 版は有意にロングショットの使用率が大きくなっている（$p < .05$）。これらの結果をまとめると、LF 版にはオリジナル映像全体の中でも映像構造・技法上の少なくとも一つの評価では、理

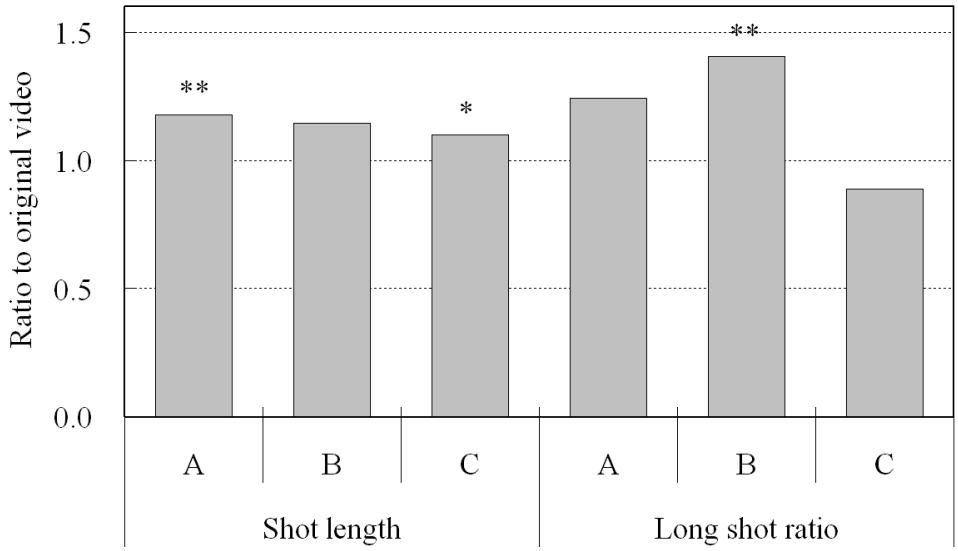

図 3.11: LF 版の映像構造（オリジナル映像との比較。** p<.05，* p<.1）

解しやすいショットが集められているといえる。

3.5　まとめ

本章では、映像理解に困難を伴わない映像区間を集めることで、短いながらも理解しやすい要約映像の生成方法を検討した。区間選択の指標には、映像理解に関わる心的負荷が低かった区間は理解しやすかったと考え、心的負荷の指標として知られる心拍変動低周波成分（LF 成分）を利用した。そして、3 タイプのオリジナル映像を 4 名の被験者に呈示することで「LF 版」の要約映像を生成した。この LF 版を評価する目的で、等時間間隔に区間選択を行うことで生成した同じ長さの「EQ 版」を対照に用い、19 名の被験者を対象に主観評価実験を行った。また、LF 版の映像構造（使用されている映像技法の利用割合）をオリジナル映像と比較した。結果をそれぞれ表 3.1 と表 3.2 にまとめた。

表 3.1 から、理解しやすさの評価結果が映像タイプ毎に異なっていることがわかる。具体的には、 映像 A（低覚醒, 中性）では映像構造のショット比の 1 項目で、映像 B（高覚醒, 快）では主観評価の映像のまとまりのよさとストーリー性及び構造評価のロングショット比の計 3 項目で、映像 C（高覚醒, 不快）では構造評価のショット比でそれぞれ理解しやすいと評価された。評価項目でばらつきがあるのは、実験刺激映像に応じて映像構成が異なり、理解しやすさの要因がそれぞれに異なるからだと考えられ

表 3.1: LF 成分を利用した理解しやすい映像要約方法の映像タイプ別適用性

Evaluation	Video A (Calm, Neutral)	Video B (Arousing, Pleasant)	Video C (Arousing, Unpleasant)
Subjective			
Cohesiveness		√	
Narrativeity		√	
Structual			
Shot length	√		√
Long shot ratio		√	
Total rating	⋆	⋆⋆	⋆

The final ratings are shown in two star rating.
No mark = not applicable, ⋆ = may applicable, ⋆⋆ = applicable

表 3.2: LF 成分を利用した理解しやすい映像要約方法の使用目的別適用性

Evaluation	Video A (Calm, Neutral)	Video B (Arousing, Pleasant)	Video C (Arousing, Unpleasant)
Personal memo	√	√	
Video message			

る。しかし、少なくとも一つの評価項目で LF 版の評価が高いこから、全体としては LF 版は比較的理解しやすいと考える。特に、（高覚醒, 快）なタイプの映像に本方法は非常に良く適しているといえる。

表 3.2 から、既視聴者が映像を思い出すきっかけに用いる映像メモ用途には、本方法は映像 A（低覚醒, 中性）と映像 B（高覚醒, 快）に適切であることがわかった。しかし、本要約映像の用途の一つである視聴した映像の第三者宛の映像メッセージとしては、本方法の有効性が限られることが示唆された。

以下に、本方法の問題点と課題を示す。

- 区間選択手順（3.2 節）が必要以上に複雑に思われる。これは、DFT を施す都合上固定長にした区間をショットと一致させるためであった。そこで、DFT の対象となる区間を必要な粒度単位でスライドするスライディングウィンドウにすることで、ショットと一致させる方がよいだろう。

- 理解しやすいことがすなわち適切な映像区間であるとは必ずしもいえないため、本方法は他方法により粗く区間選択を行った後の 2 次処理として用いられることを想定している。本章では、理解しやすい区間が選択できたかを検証することに重点を置いたためこうした処理は行わなかったが、実環境での利用に向けて、適切な 1 次処理を加えた評価実験を行うことを計画している。

- 理解のしやすさはオリジナル映像の持つ意味構造に依存するところが大きいため、特にオリジナルの段階から難解な映像では、本方法による区間選択だけでは不十分であることが示唆された。そのため、本研究ではスコープ外としていた連結手順を加味した評価実験が必要である。

第4章　印象的な要約映像

4.1　目的

本章では、1.2 節で目的に掲げた「印象的な」要約映像の生成方法を検討する。2.3 節で説明したように、印象的と感じるとは、ここでは情動の動機付けモデルで定義されるところの覚醒度が高い状態にあることを指す。そして、特定の映像区間を視聴することによって覚醒度が高くなれば、視聴者の心拍動（HR）が低下し、副交感神経活動が賦活する [16, 17, 66]。このうち副交感神経活動の賦活は、心拍変動の高周波成分（HF 成分）の増加という現象に現れる [82]。つまり、HR の低下と HF 成分の増加という二つの生理心理学的な指標を手がかりにすれば、視聴者に印象を与えた映像区間を選択できると考える。

本章ではまず、印象的な映像区間を精度よく検出する映像要約方法を確立する前段階として、これら 2 指標の効果的な組合せ方法を検討する。この目的のため、HR 低下と HF 成分増加を個別に用いてそれぞれについて要約映像を生成することで、2 指標の性質を明らかにする。以下、HR 低下を単体で使用した要約方法を「HR 方法」、この方法で生成した要約映像を「HR 版」と呼ぶ。また、HF 成分増加を単体で使用した要約方法を「HF 方法」、その要約映像を「HF 版」と呼ぶ。

続いて、得られた 2 指標に関する知見を基に、2 指標を組合せる方法を確立する。具体的には、2 指標によるそれぞれの要約映像中のショットの一致率、適合率、指標値の範囲から、組合せ時の重み付け係数を決定する。以下、2 指標を組合せて使用した要約方法を「HRHF 方法」、その要約映像を「HRHF 版」と呼ぶ。

いずれの映像要約方法についても、2.4 節で示した映像 A（低覚醒, 中性）、映像 B（高覚醒, 快）、映像 C（高覚醒, 不快）の 3 タイプの映像を用いて評価実験を行う。評価

表 4.1: 本章で生成する要約映像の種類

Digests	Method	Measures used
HR-version	HR-method	Heart Rate (HR) deceleration
HF-version	HF-method	Heart Rate Variability (HRV) High Frequency (HF) component increase
HRHF-version	HRHF-method	Both HR deceleration and HF component increase
Best-version	–	Manually-generated (subjective “best” selection)
Rnd-version	–	Randomly-generated (worst selection)

基準には、2.5 節で示した評価基準 1（適合率）を用いる。適合率算出のベースラインには、被験者がオリジナル映像から主観的に選択した区間を連結して生成した要約映像を用いる。以下、主観的には最良なこの要約映像を「Best 版」と呼ぶ。更に、HRHF 方法では評価基準 2（主観評価）も加えて評価を行う。この主観評価実験では、要約としては最も適切でない要約映像として、ランダムな選択による「Rnd 版」を先の Best 版と共に対照に用いる。本章で生成する要約映像（5 種類）を表 4.1 に示す。

以下、4.2 節で映像要約方法を説明し、4.3 節で 2 指標を単独で利用したときの方法の特性の調査を、4.4 節で 2 指標の統合とその評価を行う。最後に、本章で得られた知見を 4.5 節にまとめて示す。

4.2　映像要約方法

本提案方法を、1.3.2 節で示した一般手順に沿って示す。印象的な映像区間の登場がイベント的であることを反映し、要約映像の形式は動画形式／ハイライト型になる。要約映像の区間単位にはショットを用いる。区間選択の指標には、前述のように心拍動の低下と心拍変動高周波成分の増加を組合せて用いる。区間短縮及び連結は本研究のスコープ外とし、選択したショットを、特に処理を施さずオリジナル映像の時系列順に連結する。

区間分割

オリジナル映像を、既存のカット点検出技術（1.3.2 節）を利用してショット単位に分割する。以下、各ショットを S_s（$s = [0, \mathcal{S} - 1]$: 但し、$\mathcal{S}$ は全ショット数）と記す。

区間選択

組合せることを考えると、心拍動と心拍変動高周波成分の間で指標値の形式が揃っていることが好ましい。そこで、各視聴者から得られる 2 指標は、共に 0 ＝非覚醒的、1 ＝覚醒的というバイナリのスコアに変換する。HR 値や HF 成分値の視聴者別標準得点から求めるショット内の平均値や減少率・増加率を使用しないのは、映像区間への指標値の割り当てを、インターネットでポピュラーになっている映画などのレコメンデーションサイトやファンサイトの投票方式に似たやり方で扱いたいからである。またこのやり方ならば、全編視聴を通じての指標の全体としての変動傾向に対処しやすいと判断したからである（HR の上昇又は下降傾向については本田ら [29] を参照）。

視聴者の心拍動と心拍変動高周波成分からそれぞれ得た 0/1 のスコアは区間（ショット）に割り当ててから、まず指標別に全視聴者について合算する。この個別の合算スコアをそれぞれ単体で用いたときの要約映像が、先に説明した HR 版と HF 版になる。これらは、本研究でそれぞれの特性を調査するために生成した中間的なものである。そして、2 指標から個別に得た合算スコアに適切な重み付け係数を乗じた上で加算することで、トータルなスコアを求める（HRHF 方法）。このスコアに基づいて生成された要約映像が、本研究で目的としている印象的な要約映像である（HRHF 版）。

心拍動低下に基づく区間選択

心拍動低下に基づく区間選択手順（HR 方法）を図 4.1 に示す。

最初に、視聴者 a（$a = [0, \mathcal{A} - 1] : \mathcal{A}$ は全視聴者数）の未処理の心拍動データから、直前の値の $\pm 50\%$ を超える値をアーティファクトとして除外する。このとき、視聴者の体動情報のような外部データがあれば、それも利用してアーティファクトを除く（図

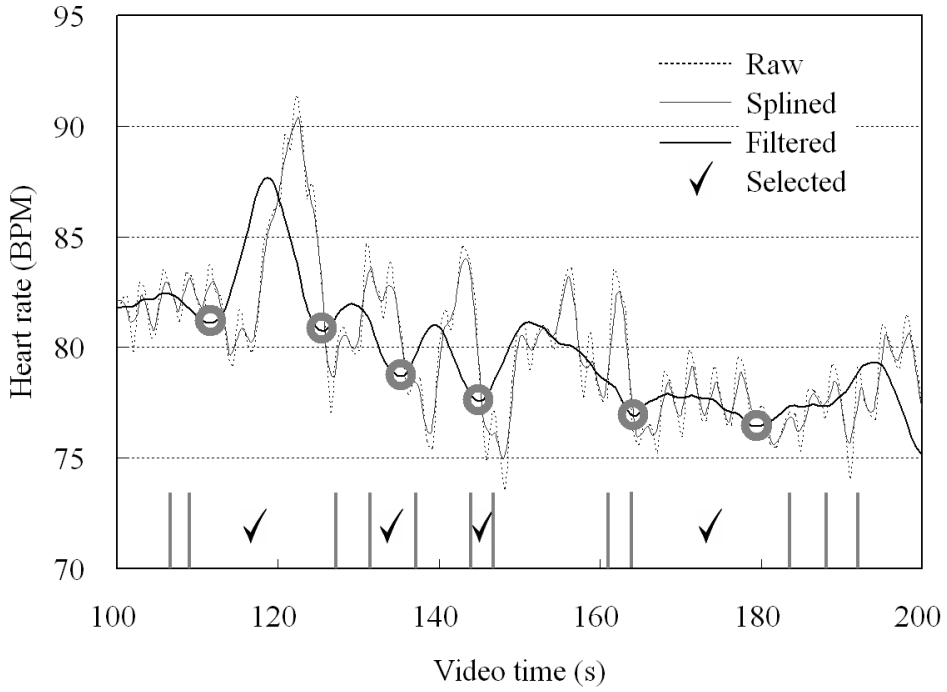

図 4.1: HR 低下を基にした区間選択方法（映像 B より）– 横軸上の長目盛はカット点を、√ は選択されたショットをそれぞれ示す

$Shots$	S_0	S_1	S_2	...	S_s	...	$S_{\mathcal{S}-1}$
$HRS_{*,0}$	0	1	1	...	1	...	0
$HRS_{*,1}$	0	0	1	...	0	...	0
⋮				⋱			
$HRS_{*,\mathcal{A}-1}$	0	1	0	...	1	...	0
HRS_s	0.3	0.9	0.8	...	0.2	...	0.1

図 4.2: 各視聴者の心拍動（HR）カーブよりショットに割り当てられた 0/1 スコアの例

中 Raw)。その後、スプライン関数を用いてリサンプリング周波数 f でリサンプルすることで HR （単位: BPM）を値とした等時系列 HR データ $hr_{i,a}$ を得る（Splined)。但し、i はサンプル点である（時刻に直せば $t = i/f$)。

次に、$hr_{i,a}$ に以下に示す 6 秒幅の移動平均処理（2.3.1 節）を施すことで、サンプル点 i における被験者 a の HR 値 $HR_{i,a}$ を得る（図中、Filtered)。

$$HR_{i,a} = \frac{1}{6f+1} \sum_{j=0}^{6f} hr_{i+j,a}$$

このフィタリング処理により滑らかな HR カーブが得られる。続いて、このカーブの谷（図中、○でマークされた箇所）が含まれているショットに対し、スコア 1 を割り当てる（√）。但し、微小な変動が谷として検出されないよう、谷判定には一定の閾値を設ける。ショットに複数の谷が含まれていても、そのショットに割り当てるスコアは 1 とする。谷と一致しないショットにはスコア 0 を割り当てる。

以上の処理を全視聴者について行えば、$\mathcal{A}$ 人分のショット単位のスコア $HRS_{s,a}$ が図 4.2 のように得られるので、そこからショット単位に全視聴者について平均を取り、ショットの HR による総合スコア HRS_s を得る。すなわち、

$$HRS_s = \frac{1}{\mathcal{A}} \sum_{a=0}^{\mathcal{A}-1} HRS_{s,a}$$

最後に、この HRS_s が大きい順に選択したショットをオリジナル映像の時系列順に連結することで、HR 版の要約映像を生成する。

心拍変動高周波成分に基づく区間選択

心拍変動高周波成分に基づいた区間選択手順（HF 方法）を図 4.3 に示す。

前処理は、HR 方法と同じである。すなわち、視聴者 a（$a = [0, \mathcal{A}-1]$）の未処理の RR 間隔データに対し、アーティファクト除去（図中、Raw）とリサンプリング周波数 f でのスプライン関数による等時系列化（Splined）を行う。但しここでは、HR の単位に BPM ではなく RR 間隔時間（単位: 秒）を用いる。このデータを $rr_{i,a}$ とする（i はサンプル点）。

次に、$rr_{i,a}$ から、視聴者 a の時刻 t における HF 成分値 $hf_{t,a}$ を得る（HF power）。これには、2.3.2 節で説明したように、t（$= i/f$）を中点とする幅 L のスライディングウィンドウに対して、離散フーリエ変換を施す。なお、この方法では、オリジナル映像の始点から $L/2$ 秒までと、終点の手前 $L/2$ 秒以降で、$hf_{t,a}$ 値に視聴前後のデータが含まれることになるが、予備実験では不都合はみられなかった。

続いて、$hf_{t,a}$ に対し、2.3.2 節で説明したように 4 秒幅の移動平均フィルタをかけることで、時刻 t の HF 値 $HF_{t,a}$ を得る（Filtered）。処理の結果、滑らかなカーブが得

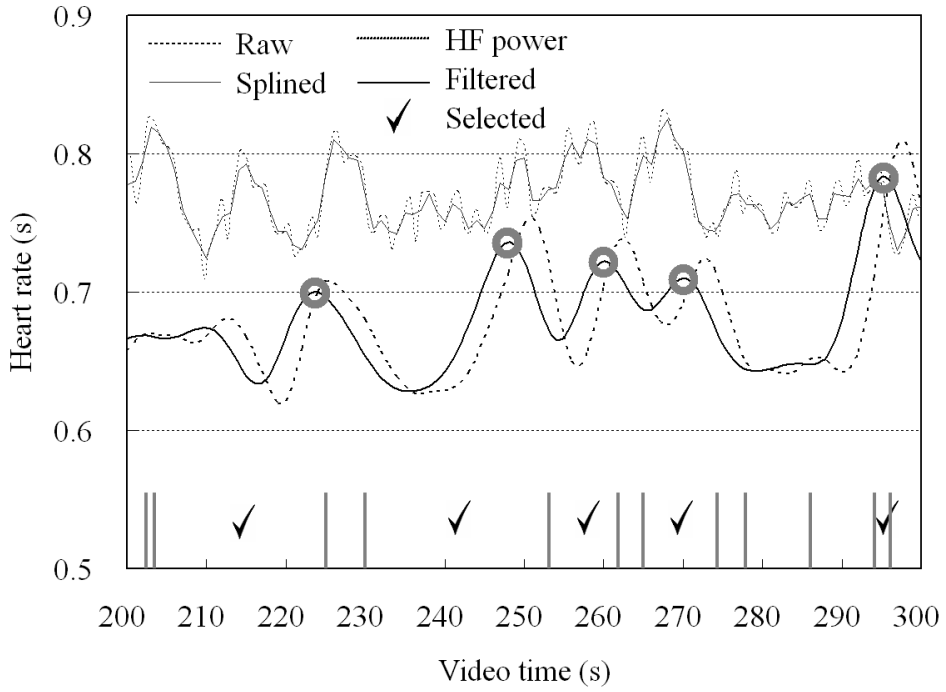

図 4.3: HF 成分増加を基にした区間選択方法（映像 B より）– 横軸上の長目盛はカット点を、√ は選択されたショットをそれぞれ示す。HF 成分値（HF power）及び移動平均フィルタ適用後（Filtered）のグラフは、見やすいように値を補正してある。

$Shots$	S_0	S_1	S_2	...	S_s	...	$S_{\mathcal{S}-1}$
$HFS_{*,0}$	1	0	1	...	0	...	1
$HFS_{*,1}$	0	1	1	...	1	...	0
⋮				⋱			
$HFS_{*,\mathcal{A}-1}$	1	1	1	...	0	...	1
HFS_s	0.4	0.9	0.5	...	0.3	...	0.2

図 4.4: 各視聴者の心拍変動高周波成分（HF 成分）カーブよりショットに割り当てられた 0/1 スコアの例

られるので、このカーブの山（図中○でマークされた箇所）が含まれているショットに対し、スコア 1 を割り当てる（√）。但し、HR 方法同様、微小な変動が谷として検出されないように谷判定に一定の閾値を設ける。ショットに複数の山が含まれていても、そのショットに割り当てるスコアは 1 とする。

以上の処理を全視聴者について行えば、全 $\mathcal{A}$ 人分のショット単位のスコア $HFS_{s,a}$ が図 4.4 のように得られるので、そこからショット単位に全視聴者について平均を取り、

$Shots$	S_0	S_1	S_2	...	S_s	...	$S_{\mathcal{S}-1}$
HRS_s	0.3	0.9	0.7	...	0.2	...	0.1
HFS_s	0.4	0.9	0.5	...	0.3	...	0.2
$HRHFS_s$	0.35	0.90	0.60	...	0.25	...	0.15

図 4.5: HR 方法と HF 方法からそれぞれ個別に得たスコア（HRS_s、HRS_s）にそれぞれ重み付け係数を乗じて合算したスコアの例（この例では $w_{HR} = w_{HF} = 1$ としている）

ショットの総合スコア HFS_s を得る。すなわち、

$$HFS_s = \frac{1}{\mathcal{A}} \sum_{a=0}^{\mathcal{A}-1} HFS_{s,a}$$

最後に、この HFS_s の大きい順に選択したショットをオリジナル映像の時系列順に連結することで、HF 版の要約映像を生成する。

2 指標の組合せに基づく区間選択

ここまでで、全 $\mathcal{A}$ 人の視聴者から図 4.5 のようなスコア表が得られるので、これらのスコアを組合せる。多くの方法が考えられるが、ここでは Hanjalic らのように重み付け係数を用いて指標を加算する方法を採用する [24]。すなわち、以下のように行う。

$$HRHFS_s = \frac{w_{HR}}{w_{HR} + w_{HF}} HRS_s + \frac{w_{HF}}{w_{HR} + w_{HF}} HFS_s$$

但し、$HRHFS_s$ は総合スコア、HRS_s は HR 方法から得たスコア、HFS_s は HF 方法から得たスコアである。また、w_{HR} は HR 方法によるスコアに対する、w_{HF} は HF 方法によるスコアに対するそれぞれ重み付け係数である。具体的な値については、4.4 節 で検討する。

最後に、この $HRHFS_s$ の大きい順に選択したショットをオリジナル映像の時系列順に連結することで 本提案方法による最終的な要約映像（HRHF 版）を生成する。

4.3　各指標の特性評価

上記で説明した、心拍動を指標とした HR 方法と心拍動高周波成分を指標とした HF 方法を個別に用いて要約映像を生成することにより、それぞれの方法の特性を評価する。本実験の目的は次の 2 点である。

- 全被験者のスコアから 4.2 節の要領で HR 版と HF 版の要約映像をそれぞれ生成し、その適合率を他の生理心理学的指標に基づく方法と比較する。
- 各被験者の指標を基に個人版の要約映像を生成し、その適合率を 1）映像タイプ別、2）被験者別、に集計することで、2 方法における映像タイプ及び個人差の影響を調査する。

刺激映像には、2.4 節で説明した 3 タイプの映像を用いた。適合率算出のベースラインとなる要約映像には、以下に説明する被験者に主観的に選択させた区間を基に生成した要約映像（Best 版）を用いた。

4.3.1　評価実験手順

まず、被験者にオリジナル映像（A、B、C）を呈示し、全被験者のスコアから HR 版と HF 版を生成した。続いて、オリジナル映像を再度呈示することでベースラインとなる Best 版を生成し、適合率を求めた。最後に、得られたデータを基に被験者毎の適合率を求めた。

HR 版と HF 版の生成

実験手順は 3.3.1 節と同じで、被験者 4 名に対しオリジナル映像（A、B、C）をランダムな順で呈示し、その間の RR 間隔データを取得した。スプライン関数のリサンプリング周波数 f は、HR 方法と HF 方法で共に 10 Hz とした。HF 方法でのスライディングウィンドウ幅は $L = 20$ 秒とし、スライディングステップは 1 秒間隔で行った。そして、選択ショット数を 10 ショットとして、心拍動低下と心拍変動高周波成分増加を

ID	時刻	時間	ショット内容	印象	驚き興奮	重要
1	00:00.0	2.6	全景カメラ1（C1)、ドイツ(独)側ゴール向き。ブラジル（ブ)、手前サイドからスローイン。			
2	00:02.6	4.0	ブ6番（カルロス）を背中からアップ。9番（ロナウド）にスローイン～カメラ、9番へ～ボールを6番へ戻すが、手前サイド際で独22番（トルシュテン）に取られる。			
3	00:06.6	16.0	C1、やや中央向き。ボール独側。コーナー側中央にボールを戻し、手前側脇線に沿って攻めるが、中央付近でチャージされ、ホイッスル。ボールはアウト。			
4	00:22.6	2.0	独19番（シュナイダー）とブ4番（ロケジュニオール)、アップ（バストショット）ですれ違う～カメラ、ブ4番の背中を追う。			
5	00:24.6	2.2	独7番（ノイビル）を側面から上半身ショット			
6	00:26.8	3.6	選手交代用紙を持って移動する審判A のバストショット。背後に独監督。			

図 4.6: Best 版生成実験用記録用紙例（映像 B、抜粋）

基にしたスコア（0/1）をそれぞれ全被験者について合算することで、HR 版と HF 版の要約映像を生成した。

Best 版の生成

評価対照用の Best 版の生成手順は次の通りである。

まず、上記の要約映像生成と同じ被験者に同じオリジナル映像を再度呈示し、印象的、驚きや興奮を覚えた、重要と考えられたショットをそれぞれ 10 ショット選択させた。呈示環境および呈示手順は 3.3.1 節と同じだが、刺激映像の呈示順序は先の生成実験と異なるようにした。選択はショット単位に口頭で行わせ、実験者がショットの簡単な説明を記した記録用紙（図 4.6）に記録した。但し、選択したショットの数は呈示中には被験者に報告していない。映像呈示は必要ならば被験者の要求若しくは実験者の判断で実験者により一時停止したが、早送りや巻き戻しは認めなかった。呈示後、被験者に記録用紙を示し、それぞれの評価項目で 10 ショットが得られなければ更に呈示を行って追加をさせた。なお、1 回目以降は被験者の要求に応じて早送りや巻き戻しをしてもよいとした。選択が 10 ショットより多ければ、被験者に記録用紙を検討させることで 10 ショットまで削減させた。そして、4 名の被験者が選択したショットにそれぞれスコア 1 を割り当てた上でショット毎にスコアを集計し、総スコアが高いショットの順に 10 ショットを抽出したものを Best 版とした。

実験後、HR 版及び HF 版に含まれている 10 のショットの中で Best 版の 10 ショットに含まれているショットをそれぞれカウントし、適合率を算出した。なお、オリジナ

ル映像からランダムに 10 ショットを抽出したときの適合率は、3 タイプの映像で平均 17% 程度になる。

映像タイプと視聴者の影響調査

続いて、被験者別に適合率を算出することで、映像タイプ間や被験者間の違いを調査した。被験者別適合率は、個々の被験者の HR 値及び HF 成分値を基に選択した 10 ショットの個人版の HR 版及び HF 版の要約映像と、その被験者の主観的選択を用いて得たスコアを基にした 10 ショットの個人版の Best 版要約映像から求めた。このとき、スコア 1 が割り振られたショット、すなわち HR 値及び HF 成分値の描くカーブの谷／山の数が 10 とは限らないため、スコアではショットの選別ができない。そこで、ここではカーブの極値（HR なら谷、HF なら山）とその直前の極値（HR なら前の山、HF なら前の谷）の差の絶対値を用いて 10 ショットを選択した。1 ショット中に複数の極値がある場合は、差分値の大きい方を指標値に用いた。そして、こうして得た被験者人数分の適合率は、映像タイプ別／被験者別に集計した。

4.3.2　結果と考察

全体の適合率と、映像タイプと視聴者の影響調査の結果を以下に示す。

全体の適合率

HR 版と HF 版の適合率を、映像タイプ別にそれぞれ図 4.7 と図 4.8 に示す。適合率の全体平均 37% と 33% は、いずれもオリジナル映像からランダムに選択した 10 ショット中に Best 版のショットが含まれている確率（適合率）である 17% より有意に大きい（$p < .05$）。

この 33% – 37%という適合率は、他の生理心理学的指標を用いた情動分類手段の平均的な適合率である約 30%（表 2.3）と同程度である。したがって、本 HR 方法及び HF 方法の区間検出力は、他の生理心理学的指標と同等程度であるといえる。但し、本

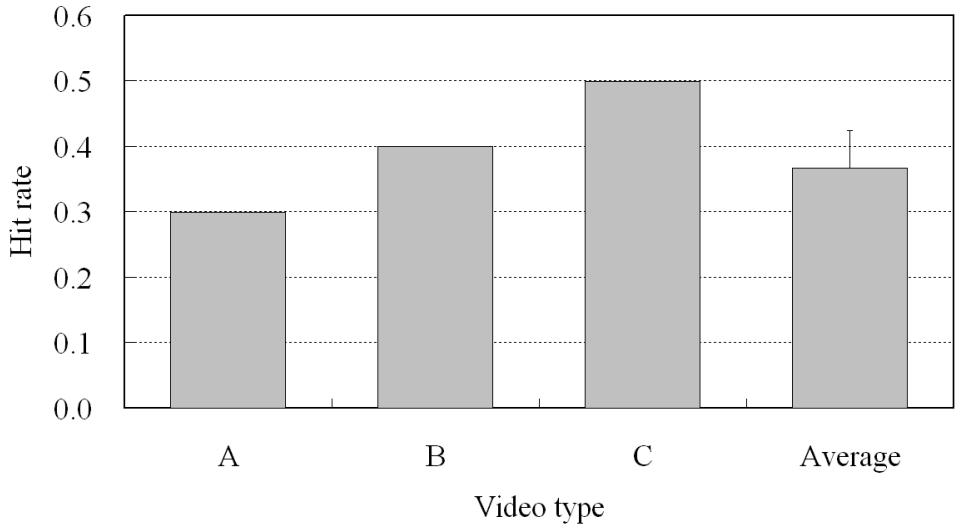

図 4.7: HR 版の適合率

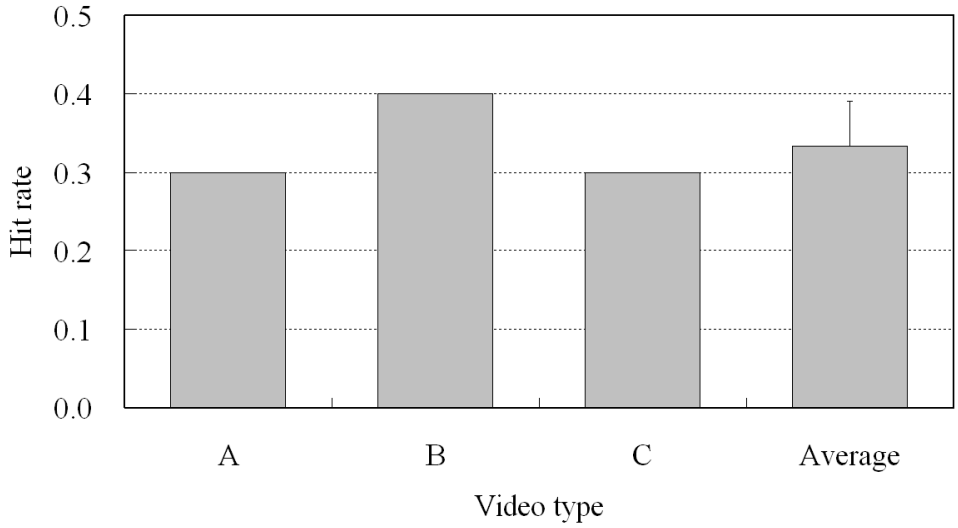

図 4.8: HF 版の適合率

方法は無拘束に生体データを取得できるというメリットがあるので、他の生理心理学的指標を用いた映像要約方法よりも実用上有利であるといえる。

HR 低下や HF 成分増加の検出には、本方法が採用している 0/1 のバイナリスコア（投票方式）以外にも、ショット内の平均値や値の減少率・増加率を用いる方法が考えられる。そこで、本方法をこれらの方法と比較する目的で、HR 低下について、それぞれの方法で生成した要約映像の適合率を比較した。どちらの場合でも、HR 値は被験者毎に $[0, 1]$ の値となるように正規化した。減少率については、i 時点の指標値を

$HR_{i,a} - HR_{i-1,a}$ とし、区間中でこの値が負のものについてのみショット単位で平均した値を用いた。その結果、適合率は平均方法では 10 ± 10%、減少率方法では 24 ± 6% であった。このことにより、本方法の用いるスコア式は区間選択方法として少なくとも平均や減少率よりは妥当であることが確認された。

映像タイプと視聴者の影響

HR 方法において、被験者単位に生成した個人版要約映像を基に算出した映像タイプ別の適合率平均を図 4.9 に示す。この結果から、映像タイプ別には特に有意差はないため、本方法はどのタイプにも適用可能と考えることができる。

同様に、適合率を被験者別にまとめた結果を図 4.10 に示す。被験者 1 と被験者 2 の間に有意差が見られるが（$p < .05$)、他の被験者間ではみられない。ここから、視聴者によっては本方法が適さない可能性のあることが示唆される。この結果は、覚醒による HR 低下は一般的な傾向としては正しいが、およそ 25% はそういう傾向を示さないという知見 [6] と一致する。そうならば、本方法は視聴者数が少ない場合、HR 低下を示さない視聴者のスコアが過大に集計スコアに反映してしまうことで、覚醒的＝印象的なショットを選択する能力が低下することになる。この問題は、一定人数以上のデータが集まらなければ要約映像は生成しない、既存のレコメンデーションサイトやファンサイトのように現在の投票数を示すことで情報の確度を示すようなインタフェースを提供するといった、実装時の配慮で解決できると考える。なお、Web 上の映画ファンサイトを調査したところ、ポピュラーな映画の約 89% が 6 名以上からの投票を得ているので[1]、心拍活動ベースの本方法が仮に同程度の投票を得ることができるとしたら、大半はカバーできるとしてよいだろう。

次に、HF 方法において、被験者単位に生成した個人版要約映像を基に算出した映像タイプ別の適合率平均を図 4.11 に示す。この結果から、本方法は映像 B（高覚醒, 快）のタイプでは特に効果的であるが、映像 A（低覚醒, 中性）と映像 C（高覚醒, 不快）

[1] http://allcinema.net/から、1）古今のアカデミー作品賞受賞作品、2）AFI （American Film Institute）の「アメリカ映画 100 年ベスト 100」、3）キネマ旬報「オールタイムベスト ベスト 100 外国映画編（1999 年度版）」から重複なく選んだ映画 235 本を基に調査した。

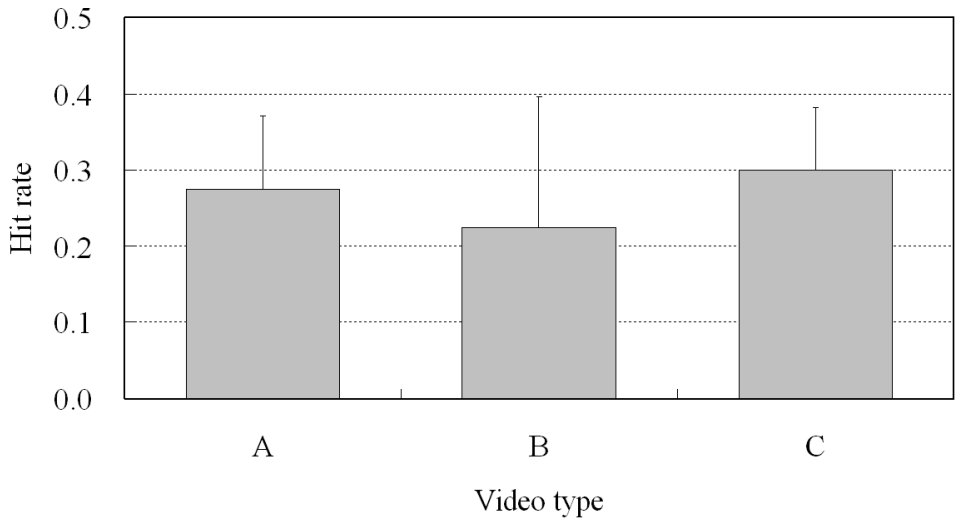

図 4.9: 個人版の HR 版の映像タイプ別適合率

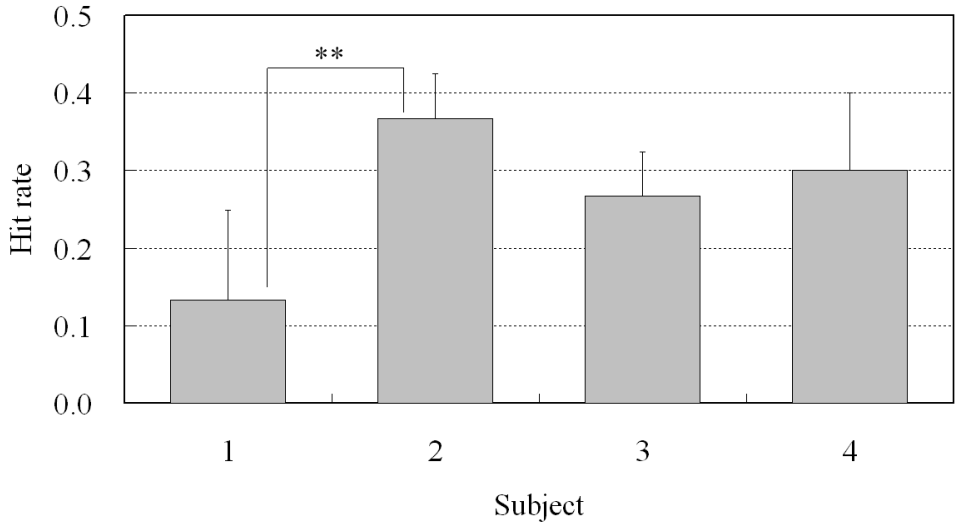

図 4.10: 個人版の HR 版の被験者別適合率（** $p < .05$）

なタイプでは区間検出力が弱いことが示唆された（A-B 間に有意差（$p < .05$）、B-C 間に有意傾向（$p < .1$）が見られた。A-C 間には有意差はみられなかった）。ここから、HF 方法には、HR 方法と異なり映像タイプ依存性があることがわかった。

同様に被験者別にまとめた結果を図 4.12 に示す。被験者間には特に有意差はみられないため、HF 方法の視聴者依存性は少ないことが示唆された。

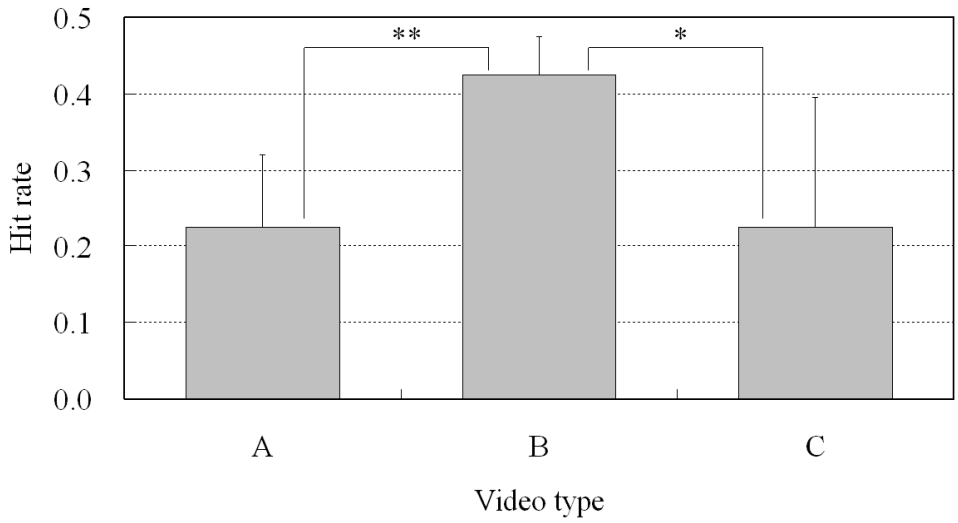

図 4.11: 個人版の HF 版の映像タイプ別適合率（** $p < .05$、* $p < .1$）

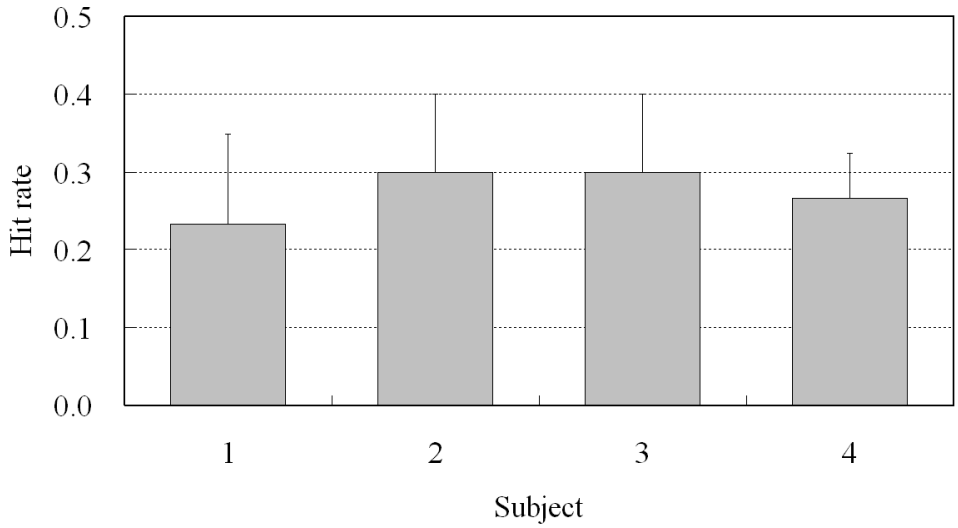

図 4.12: 個人版の HF 版の被験者別適合率

4.4　指標の複合化による精度向上

以上で、心拍動（HR）と心拍動高周波成分（HF 成分）をそれぞれ個別に利用したときの映像要約の特性が明らかになったので、続いて、これら 2 指標を組合せた映像要約方法を確立する。2 指標を組合せるのは、生理心理学的指標を用いた情動分類メカニズムでは複数の指標の組合せで精度が向上することが一般的に知られており [68]、本方法でもこれが期待できるからである。

しかしその前に、1）同じ心拍活動起因の指標を組合せることに意味があるか、2）あるならば、どのように組合せるのが効果的であるか、を明らかにする必要がある。本節ではまず組合せの意義と方法について、既存の知見と 4.3 節の結果を用いて検討する。続いて、HR 方法と HF 方法から得たスコアを組合せること（HRHF 方法）により要約映像（HRHF 版）を生成し、適合率と主観評価を通じて評価を行う。

4.4.1 組合せ方法の検討

まず、既存の知見から、心拍動と心拍変動高周波成分というどちらも心拍活動起因である指標を組合せることに意味があるかを検討する。

心拍の挙動を支配する自律神経の活動は一般に、副交感神経（PNS）が賦活すると交感神経（SNS）が後退するというように相反的な挙動を示す。このような挙動は、例えば心拍を一定に保つ恒常機能でみられる（e.g. 過度な心拍動低下は SNS の賦活と PNS の後退により上昇に転じられる）。この場合、PNS の増加は心拍動低下と直接関係する。そうならば、HR か HF 成分の一方の指標だけ利用すれば十分であり、仮に両者を組合せて用いてもノイズや計算誤差の抑制程度の効果しか期待できないため、精度向上はさほど望めないことになる。

しかし、心拍動の変化が行動に関わる高次の脳活動（特に辺縁系や前脳）による場合、自律神経は必ずしも相反的な挙動を示すわけではない。その挙動は、SNS と PNS が互いに方向性が異なる挙動を示す相反的（reciprocity）、双方が共に賦活若しくは後退する共動的（coactivity）、互いに無関係な挙動を示す非共役的（uncoupled）と多様である [10, 11, 16, 66]。本研究で取り扱う覚醒状態に伴う心拍動低下も、2.3 節で説明したように、SNS と PNS の双方の賦活（但し PNS が優位）という共動的な挙動に基づいて生起すると説明されている [16, 17, 66]。そのため、HR 低下だけに依存すると、例えば一定な PNS と抑制された SNS という覚醒的ではないと考えられる状態も覚醒的と検出してしまうことになる。また逆に HF 成分（PNS）の増加だけに基づくと、PNS 以上に SNS が賦活することによる心拍動上昇のような状態も検出してしまう。こうしたことを考慮すると、HR と HF 成分を組合せることには意義があると考

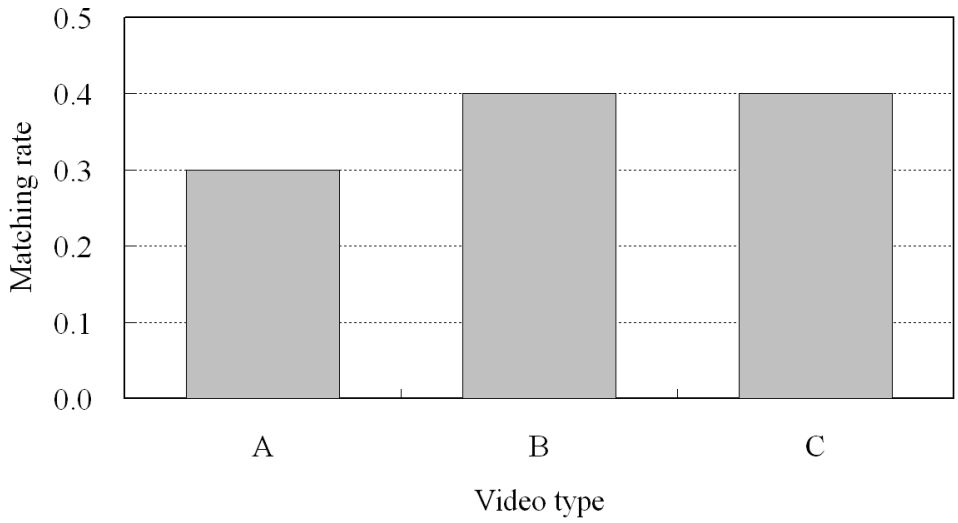

図 4.13: HR 版と HF 版のショットの一致率

える。

上記のように、HR と HF 成分が必ずしも同期的でないとするならば、HR 方法と HF 方法が選択したショットはそれぞれ微妙に異なるはずである。このことを確認するため、4.3 節で生成した HR 版と HF 版が同一のショットを選択した割合を調査した。結果を図 4.13 に示す。一致する割合は 30% – 40%（10 ショット中、3、4 ショットが一致）であり、一部同期的であるがそうでもない部分もあるという結果となった。ここから、HR と HF 成分という二つの心拍活動起因の指標を組合せることによる効果は期待できると考える。

続いて、組合せの方法を検討する。区間選択は、4.2 節で説明したように、HR 方法で得たショットのスコアと HF 方法で得たショットのスコアにそれぞれ適切な重み付係数（w_{HR}、w_{HF}）を乗じて加算することで得た総合スコアを基に行うが、ここで重要なのはこれら重み付け係数の決定である。ここでは、Hanjalic のように指標値の範囲を基にすると共に、HR 版と HF 版の区間選択の適合率を基に決定する。後者は、精度の高い方法により重み付けをすることで、区間検出精度を向上させるための措置である。

まず指標値の範囲だが、これは視聴者あたりのスコア 1 の数、すなわち映像呈示中の HR 及び HF カーブの極小点（谷）若しくは極大点（山）の数に相当する。図 4.14 に、4.3 節の評価実験で得た谷と山の平均数を示す。山及び谷の数は同程度（$p < .05$ で有

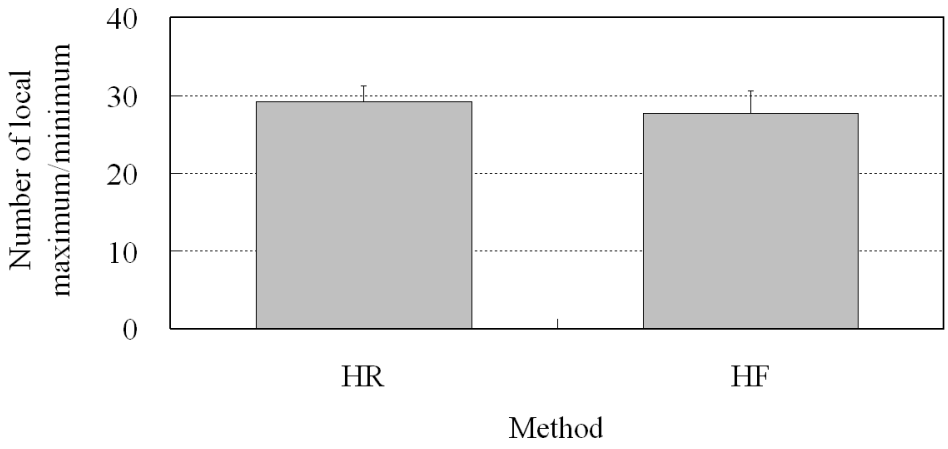

図 4.14: HR 方法と HF 方法による刺激映像中の極大値／極小値の数

意差なし）であるので、この観点からは w_{HR} と w_{HF} は等しく取ればよいということがわかる。適合率については、HR 版と HF 版の全映像タイプでの平均適合率は 37% と 33% であり、その間に有意な差は認められない。そこで、重み付け係数は等値とする。すなわち、

$$w_{HR} = w_{HF} = 1$$

とする。

4.4.2 評価実験手順

上記で決定した重み付け値を用いて、2.4 節で説明した 3 タイプの映像（A、B、C）から要約映像（HRHF 版）を生成し、その妥当性を適合率と主観評価実験で評価した。

適合率

HRHF 版を、4.3 節と同じく 10 ショットを選択することで生成した。そして、4.3 節で生成した被験者の主観的選択による要約映像（Best 版）をベースラインに用いて適合率を算出した。

表 4.2: HRHF 版主観評価実験で用いた要約映像の時間長（単位：秒）

Method	A	B	C
HRHF	95.3	110.7	100.7
Best	84.3	99.9	82.8
Rnd	94.2	100.4	81.2

主観評価

続いて、HRHF 版に印象的なショットが有効に集められているかを主観的に評価した。この実験では、10 名の被験者にオリジナル映像を呈示し、その直後に HRHF 版、Best 版、そして新たに比較対照用に用意した「Rnd 版」を呈示し、それぞれについて主観評価を行わせた。この Rnd 版は、オリジナル映像からランダムに 10 ショットを選択して時系列順に連結することで生成した要約映像である。但し、ランダムに 10 ショットを選んで Best 版のショットと一致する確率（17%）を基に、一つか二つは Best 版中のショットが含まれるように調整してある。本実験で用いた HRHF 版、Best 版、Rnd 版の映像タイプ別の時間長を表 4.2 に示す。

評価項目は次の 4 項目で、いずれについてもオリジナル映像を視聴したときに項目に合致するショットが要約映像に含まれている割合を多い（5）から少ない（1）の範囲で 5 件法で回答させた。

- 印象的だったショット
- 興味関心を惹いたショット
- 注意を喚起したショット
- 重要であったと感じられたショット

Best 版は主観的には最も完全な、Rnd 版は最も適切でない要約映像となるので、HRHF 版が Best 版と同程度若しくは Best 版と Rnd 版の中間と評価されれば本方法が妥当であると判断した。すなわち、印象的なショットが多いと評価されればより印象

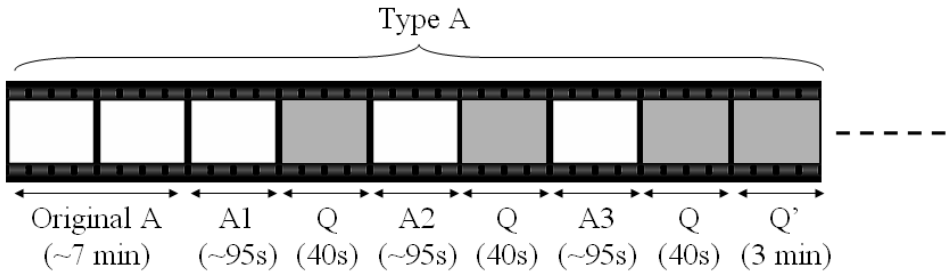

図 4.15: HRHF 版主観評価実験プロトコル – 図は映像タイプ A のみだが、他も同じ構成である。映像タイプの呈示順序は被験者毎にランダムに入れ替えた。A1、A2、A3 は Best 版、HRHF 版、Rnd 版の要約映像で、呈示順序は映像タイプ毎にランダムに入れ替えた。

的であることから、上記評価項目の評点が Rnd $<$ HRHF $\leq$ Best となる映像タイプが本方法の適用可能範囲となる。

実験プロトコルを図 4.15 に示す。それぞれの映像タイプについて、まずオリジナル映像（図中、Original A）を呈示し、続いて 3 種類の要約映像（A1、A2、A3 は Best、HRHF、Rnd 版のいずれか）を 40 秒の回答期間（Q）をはさんで呈示した。回答期間には、ディスプレイ上に上記 4 項目を表示し、被験者に口頭で評点を回答させた。最後に、3 分間のアンケート（Q'）を行った。3 種類の要約映像は、映像タイプ毎にランダムに入れ替えた。この手順は映像タイプすべてについて、呈示順序を被験者毎にランダムに入れ替えて行った。呈示環境は、LF 版の要約映像生成実験（3.3.1 節）と同じだが、映像呈示中に、被験者正面に設置したビデオカメラで被験者を撮影した。被験者には参考のためであり、無視するように教示した。

4.4.3 結果と考察

適合率評価実験と、3 種類の要約映像を用いた主観評価の結果を以下に示す。

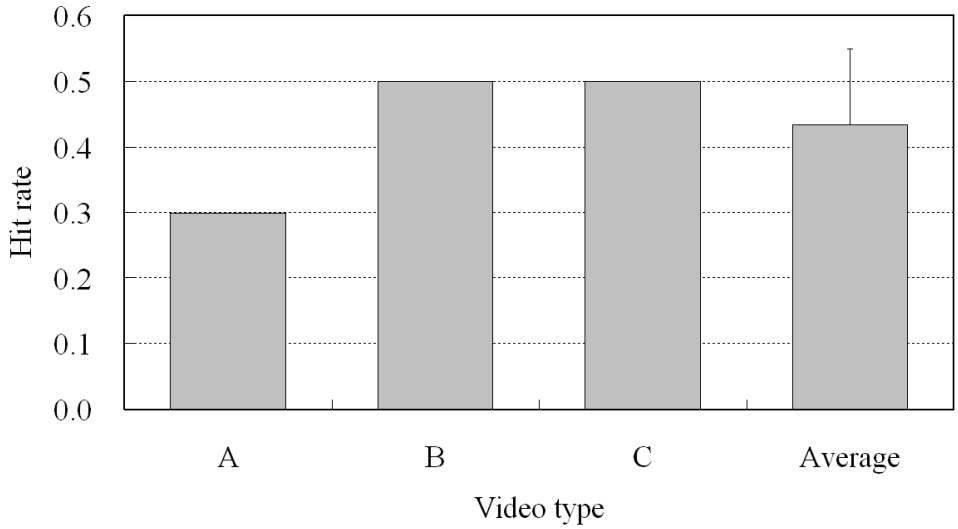

図 4.16: HRHF 版の適合率

適合率

本 HRHF 方法の適合率を図 4.16 に示す。適合率の全映像タイプ平均は 43% であり、2 種類の生理心理学的指標を組合せた高橋の結果である 25±11% [79][2] よりも高く、全体を通してみても適合率範囲（表 2.3）の高いレベルにある。このことから、本提案方法は生理心理学的指標を用いた方法としては精度的には問題はないということができる。これに加え、無拘束的な心拍センサを利用できることから生体データの収集に伴う視聴者の負担を軽減でき、また広範囲に要約映像システムを展開できるという実用上の優位性が示された。

次に、HR 方法と HF 方法を組合せることにより適合率向上が達せられたかを確認する。HR 版及び HF 版と HRHF 版の適合率には、全体平均では有意差はみられなかった。しかし、HR 版、HF 版、HRHF 版のいずれでも同じ適合率の映像 A（低覚醒, 中性）を精度向上がみられないタイプとして統計から除くと、HR 版／HF 版からの精度向上に有意差がみられた（$p < .05$）。このことから、（低覚醒, 中性）のタイプの映像を除き、本方法により区間選択の精度が上げられることがわかった。

[2]高橋の 2 指標利用情動分類メカニズムは SVM（Support Vector Machine）を用いている。

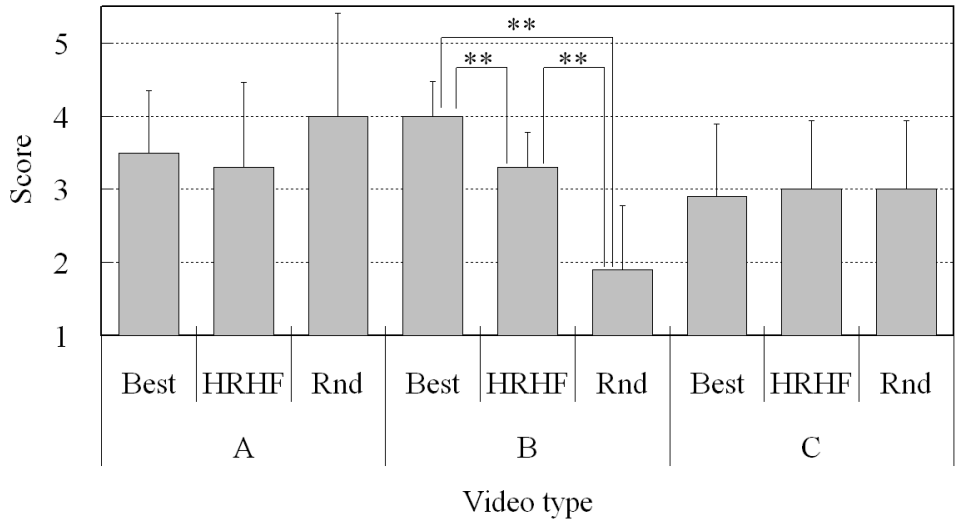

図 4.17: HRHF 版の印象の度合い（** p<.05）

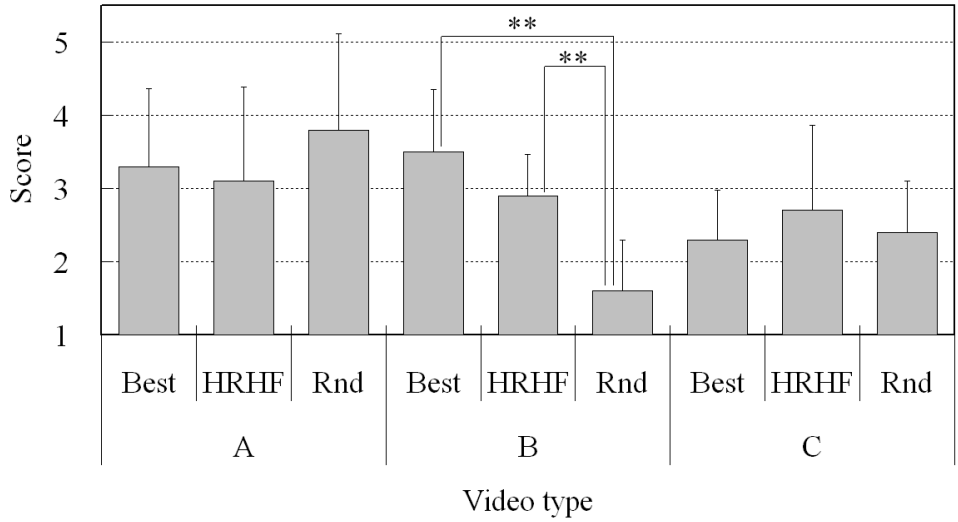

図 4.18: HRHF 版の興味関心の度合い（** p<.05）

主観評価

本方法による要約映像が視聴者に与える印象度を主観評価を通じて調査した結果を図 4.17、図 4.18、図 4.19、図 4.20 にそれぞれ示す。

映像 B（高覚醒, 快） では、いずれの評価項目でも印象的なショットを含む順が Rnd < { HRHF, Best } であり、本方法が（高覚醒, 快）な映像に適用できることが示された。また、4 項目のうち 3 項目で Best 版と HRHF 版の間に有意差がみられなかった

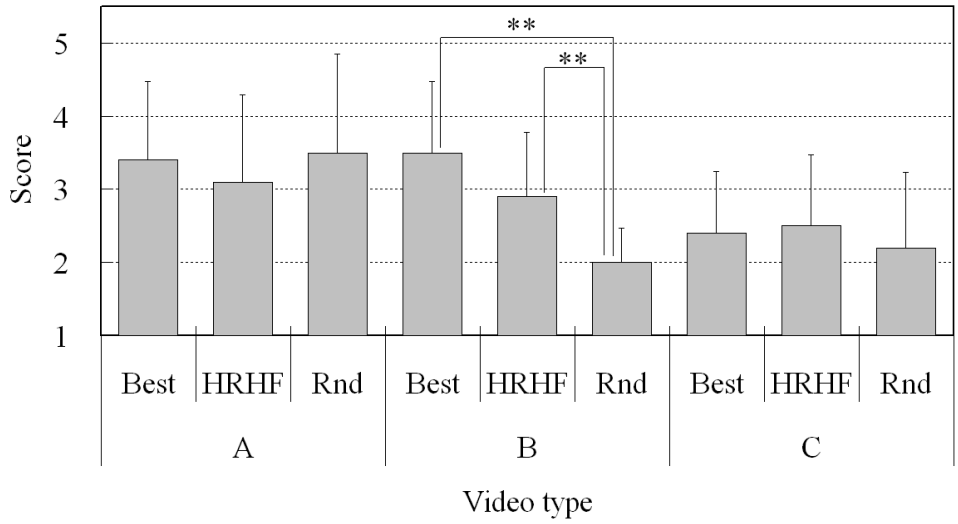

図 4.19: HRHF 版の注意の度合い（** p<.05）

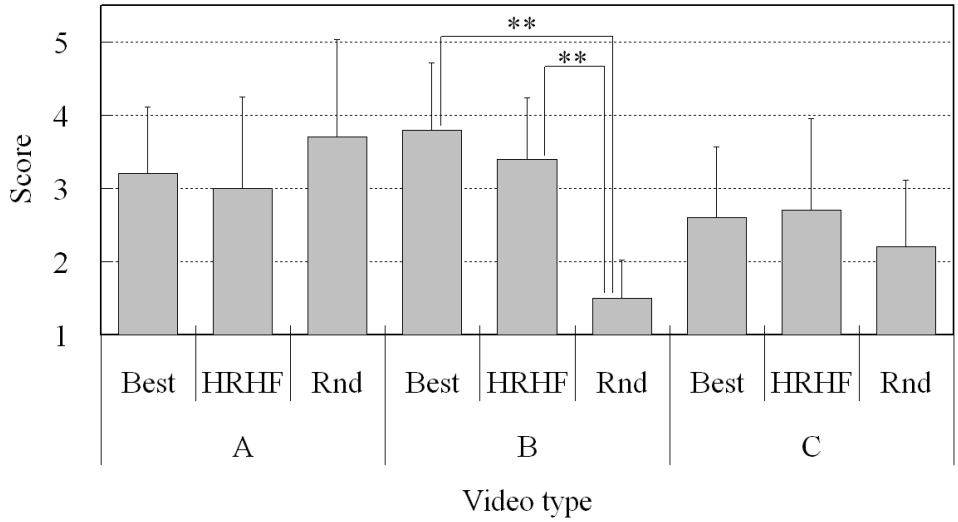

図 4.20: HRHF 版の重要さの度合い（** p<.05）

ことを考えると、映像 B のようなタイプの映像では、要約映像中に主観的に覚醒的と判断されるショットが半数以上含まれていれば実用上は問題がないと想定できる。しかし映像 A（低覚醒, 中性）と映像 C（高覚醒, 中性）では差は認められなかった。

映像 A（低覚醒, 中性）については、低覚醒な映像では HR の低下や HF 成分の増加が顕著ではないことによる精度不充分と、2 指標の組合せが上述のように効果を発揮しなかったためと考えられる。このことは、適合率が 30% と他と比較して低いことか

らも伺える。差のなさはまた、鳥の飛翔だけの似通ったショットでほとんどが構成されているというオリジナル映像の性質に起因しているとも考えられる。インタビューでも、特に印象に残っているようなショットが要約映像中に見当たらない、若しくはいずれも印象的であったという報告が得られている。印象的なショットがあったと回答した被験者でも、好みのショットはまちまちであった。こうしたことから、本方法は同じようなショットやシーンが続くような平板な映像については、スコアが視聴者毎にばらついて特定のショットを絞りきれない状態になるためにショット選択の確度が低くなることが示唆される。しかし、この問題はレコメンデーションサイトやファンサイトが毀誉褒貶の定まらない対象を扱うときに共通した問題であるので、確度情報として投票件数ヒストグラムなどを併せて提供するなどの画面設計を考慮すればよいと考える。

映像 C（高覚醒, 不快）で差が示されなかったことについては、被験者間で印象的と考えるショットが分かれたためだと考える。映像 C は 2 部構成になっており、前半部分で行われたことが後半部分で登場人物の立場を入れ変えて同じ様式で繰り返される。HRHF 版に含まれているショットは、ほとんど前半部分に属している。これは、恐らくは慣れ（habituation）[5] が要因で後半部分での HR 低下と HF 増加が前半部分ほど顕著でなくなったためと考えられる。またインタビューでも、ストーリーの繰り返しそのものが印象に残ったと述べた数名の被験者が、繰り返しを示さない要約映像では感じた印象を適切に示していないと評価している。このことから、同種の内容の繰り返しが印象形成に寄与しているオリジナル映像に対しては、本方法は繰り返しを示さないために不満足な結果をもたらすという示唆も得られた。しかし、要約映像長を長く取れば繰り返しは示されるので、要約映像長を制御するわかりやすいインタフェースを提供することで、この問題は運用上回避できると考える。但し、要約映像長を短くする場合は、シーンやシーケンスといった映像構造を把握し、区間選択を各シーン・シーケンスからまんべんなく行うような方法が必要になると考えられる。

以上から、心拍活動起因の HR と HF 成分を指標に利用した本映像要約方法は、（高覚醒, 快）なタイプの映像で有効であることがわかった。また、低覚醒な映像、ショット間に差異があまりない漫然とした映像、若しくは映像構造の繰り返しが印象形成に

寄与するタイプの映像では期待通りに動作しない可能性があることが示唆された。また、視聴者によって評価の分かれる映像には、要約結果がどの程度の確度で覚醒度を表しているかを示す補助情報や、要約映像長をユーザ側で制御できるインタフェースを用意する必要があることがわかった。更に、映像構造が印象形成に影響するタイプのオリジナル映像については、構造の把握が重要であることも示唆された。

4.5 まとめ

本章では、映像の持つ情動的な側面に着目し、視聴者にとって印象的な映像区間（ショット）を集めた要約映像の生成方法を検討した。ここで印象的であるとは、情動の動機付けモデルが定義するところの覚醒度に対応すると仮定している。そして、高覚醒な映像を視聴すると、1）心拍動（HR）が低下するという情動の動機付けモデル、2）刺激受容期には副交感神経が賦活するという防衛の段階的反応モデルと副交感神経系の活動は心拍変動高周波成分（HF 成分）に反映するという知見、に基づき、各視聴者から得たこれら 2 指標より算出したショット単位のスコアを全視聴者について合算することで映像要約を行った。

本研究では、これら 2 指標の特性を明らかにする目的で、まず 2 指標を単体で用いたとき（HR 方法と HF 方法）の適合率を調査した。そして、その結果を基に両者を組合せる方法を検討し、主観評価実験を通じて本提案方法（HRHF 方法）の効果を映像タイプ別に明らかにした。得られた知見は以下の通りである。

2 指標を単体で利用した映像要約

HR 方法及び HF 方法の特性を表 4.3 に示す。

HR 方法と HF 方法の適合率は、全映像タイプの平均でそれぞれ 37% と 33% であった。この結果は、他の心理生理学的指標を用いた情動分類手段と同程度の適合率である。これにより、心拍動及び心拍変動高周波成分をそれぞれ単体で用いた場合でも、無拘束なデータ収集が可能というメリットを考慮すると、実用度が高いことがわかった。

映像タイプに対する依存性を適合率から調査したところ、HR 方法ではどのタイプ

表 4.3: HR 方法と HF 方法の特性

Characteristics	HR-method	HF-method
Precision	37 %	33 %
Viewer dependency	Yes	No
Video-type dependency	No	Yes

でも同程度だが、HF 方法では映像 B（高覚醒, 快）では有意に効果が高く、他 2 タイプでは低いという依存性があることがわかった。

また、視聴者に対する依存性（個人差）について調査したところ、HR 方法では 1/4 の被験者で有意に適合率が低いことがわかった。これは、4 人に 1 人程度は覚醒的画像を呈示されても心拍動低下が発生しないという知見と一致し、本方法は適切な区間選択精度を保つには 4 人以上の視聴者を必要とするという実用上の指針も得られた。反面、HF 方法には視聴者依存性がないことがわかった。

2 指標を組合せて利用した映像要約

同じ心拍動起因の二つの指標を組合せることについては、HR 方法と HF 方法が異なるショットを選択する（一致率は 30% – 40%）ことから、効果があると考えられた。実際に組合せを行うと、映像 B（高覚醒，快）と映像 C（高覚醒，不快）で適合率が向上することが確認されたので、高覚醒な映像タイプでは組合せが効果的であるとわかった。

組合せ方法については、HR 版と HF 版の適合率と値範囲から、HR 方法と HF 方法からそれぞれ得たショット単位のスコアに同じ重み付け係数を乗じて加算すればよいという経験則が得られた。

本方法（HRHF 方法）の適合率は、全映像タイプの平均で 43% であった。この結果は、他の生理心理学的指標を用いた情動分類手段の適合率と比較すると高いレベルに位置しており、満足のいける結果だといえる。

実験では、本方法で生成した要約映像を、主観的な選択による「最良」な要約映像（Best 版）とランダムな選択による「最悪」な要約映像（Rnd 版）と共に呈示することで主観的に評価した。この結果、映像 B（高覚醒, 快）では、本方法はよく適用できることが示されたが、他の映像タイプについては映像内容が平板すぎる、若しくは映像構造が要約映像に反映されていないなどの理由から、適用性が限られるという結果となった。

以上の結果を、評価項目別に表 4.4 にまとめた。映像 B（高覚醒, 快）は適合率が高

表 4.4: HRHF 版の映像タイプ別適用性

Evaluation	Video A (Calm, Neutral)	Video B (Arousing, Pleasant)	Video C (Arousing, Unpleasant)
Precision	⋆	⋆⋆	⋆⋆
Combination		⋆	⋆
Subjective		⋆⋆	
Total rating		⋆⋆	⋆

The ratings are shown in two star rating.
No mark = not applicable, ⋆ = may applicable, ⋆⋆ = applicable

く、2 指標組合せの効果もあり、また主観評価の結果も Best 版と遜色のないほど高いことから、本方法が非常によく適合すると考えられる。映像 C（高覚醒, 不快）でも適合率と組合せはよいが、主観評価は低いため、本方法がおおむね利用できるという結果になった。しかし、映像 A（低覚醒, 中性）には、適合率は他の生理心理学的指標を用いた情動分類手段の平均的な値と同等程度だが、2 指標組合せの効果が得られず、また主観評価でも評価が低いために適切ではないとわかった。

今後の課題

以下に、本提案方法の問題点と今後の課題を示す。

- 印象の度合いを示す心理生理学的な指標は心拍活動以外もあるので、これらを利用することで適合率の向上を図る。例えば、視聴中に身を乗り出したり反らしたりするといった姿勢情報は、映像の印象度と大きく関係していると考えられる [36]。姿勢情報の都合のよい点は、感圧式である心拍センサ（2.1.1 節）から心拍活動データ収集と同時に無拘束に取得できる点にある。感圧式センサからはまた、呼吸数も取得できる [64, 87]。呼吸数は HF 成分同様副交感神経系の活動を示す指標であるので、HF 方法の精度向上に利用できる。更に視聴者に向けたビデオカメラを活用できれば、表情 [83]、瞬目 [36]、顔面の肌色を解析することで得られる交感神経系（SNS）の活動度 [30] も視聴者を拘束することなく収集が可能に

なる。これらはいずれも、指標の複合化による精度向上に寄与すると考える。

- 映像 C で示されたように、よい要約映像は映像構造を反映している必要がある。これについては、映像内情報を用いてシーンやシーケンスといった映像構造を把握する研究（e.g 加藤ら [38]、是津ら [43]）の成果の応用を考えている。例えば、本研究ではオリジナル全編から指標値順にショットを選択したが、まずシーン又はシーケンスに分割しておき、シーン・シーケンス内で指標値の高いショットを一つ選択するような手順を組み込むといった方法が考えられる。

- 印象の強度は覚醒度と相関するため、覚醒度の低いオリジナル映像に対しては区間選択精度が上がらないという問題がある。当然ながら、全編を通じて覚醒度は上下するものなので、本方式により中でも覚醒度の高いショットを選択できるようにはなっているはずだが、結果としては低覚醒な映像タイプ A で適合率が低くなっている。これは、HR ／ HF カーブの変動が大きくないために検出精度が上がらず、そのために区間選択が適切にできなかったことに起因していると考えられる。しかし、このことは、本研究の多様な映像やユーザニーズに対応するために各種の区間選択指標を提供していくという目的と矛盾するものではない。オリジナルの段階から覚醒的ではない映像については、覚醒度を基にした映像要約は不適切であると考えられるので、最初に覚醒的でないと別手段（e.g. 映像内情報、HR ／ HRHF カーブ全体の変動率）で判定されたオリジナル映像に対しては本方法を適用せず、例えば「観るとリラックスする」のような別の視聴経験的内容で映像要約を行うべきであろう。

第5章　結論

本論文では、視聴者の映像視聴経験を反映した映像要約方法を確立する目的で、次の2種類の要約映像を作成した。

1. 理解しやすい要約映像

2. 印象的な要約映像

本研究では、このような目的には、視聴時の認知的・情動的な心的活動を反映した生理心理学的な反応を利用することが適切であると考え、視聴者の心拍活動から映像区間選択の指標を抽出した。心拍活動には、無拘束な生体センサ技術が利用でき、視聴者の視聴経験を損なわずにデータ取得ができるという実用上の利点がある。目的1については、理解しやすい＝心的負荷が低いとし、心的負荷の指標である心拍変動低周波成分を用いた。目的2については、印象的＝覚醒的であるとし、覚醒度の指標である心拍動と心拍変動高周波成分を用いた。

第1章では、本研究の背景、目的、既存の映像要約技術における位置付けを示した。続く第2章では、視聴経験を推定する上で必要な心拍活動指標に関する基礎的な知見をまとめ、その処理方法の概略を示した。また、本提案方法の評価実験で用いる刺激映像と評価基準を示した。

第3章では、心拍変動低周波成分を用いて上記目的1の「理解しやすい」要約映像の生成方法を検討し、主観評価と映像構造を基準とした評価実験を通じて以下の知見を得た。

- 心拍変動低周波成分は、映像区間の理解しやすさを検出するのに有効である。特に、映像の覚醒度が高く、情動価が快に分類されるタイプの映像では効果的である。

- 本要約映像は、オリジナル映像の視聴者が視聴した場合は理解されやすいが、観たことのない第三者には元視聴者ほど理解しやすいものではない。このため、視聴者本人がメモ代わりに利用するという用途には適切だが、他者にメッセージ（e.g. 友人に観たものを紹介する）として呈示するという用途では効果が限定される。

- 映像の理解しやすさは映像の構造に依存する。このため、より実用的な要約映像の生成には、ショットよりも大きい意味区間であるシーンやシーケンスを考慮に入れる必要がある。

第4章では、心拍動と心拍変動高周波成分を指標に用いて上記目的2の「印象的な」要約映像を生成する方法を検討し、適合率（被験者の主観的な選択区間に対する本方法の一致率）と主観評価を基準とした評価実験を通じて以下の知見を得た。

- 心拍動と心拍変動高周波成分を組合せて印象的な映像区間を選択する本方法の適合率は、43% であった。これは、他の生理心理学的指標に基づく方法と比較すると高いレベルの精度であり、無拘束取得が可能であるというメリットを考え合せると、本方法の実用性の高さを示すものである。

- 心拍動を単体で用いた場合、視聴者によっては適切な映像区間の選択ができないことがある。そこで本方法の利用に際しては、4人以上の視聴者が必要になる。

- 心拍動と心拍変動高周波成分を組合せて利用するに際しては、重み付けをせずに加算するとよい。但し、組合せの効果は、対象となる映像のタイプが覚醒度の高いタイプに限られ、低覚醒なタイプの映像では効果が現れない。

- 印象の度合いは映像の持つ構造に影響を受けるため、シーンやシーケンスといった映像構造を考慮する必要がある。

従来の映像要約研究では、映像内の物理的な特徴量を用いて、ゴールシーンや特定のキーワードが語られた場面の検出といった、映像の客観的・説明的な観点に基づいた映像要約に主たる関心が置かれていた。こうした方法も当然ながら重要ではあるが、

要約映像の利用者のニーズはより多様であり、より魅力的な要約映像を提供するには要約方法の多様化が欠かせない。そこで、「エキサイティングな場面だけを集めた要約映像」のように、映像視聴を通じて得られる視聴経験という観点に基づいた映像要約方法が国際的にも注目を集めるようになってきた。本論文はこうした要求に対し、生理心理学的指標という人間工学的なアプローチを応用することにより、映像要約の多様化に貢献できたと考える。

今後は、提案方法の効果を高めたり用途を広げるといった課題に取り組むことを計画している。具体的には、以下のような課題がある。

- 本提案方法は、無拘束型の感圧センサを利用した心拍活動の取得を想定しているが、感圧センサからは心拍活動以外にも、身を乗り出す・反らすといった体動情報や副交感神経活動の指標を抽出できる呼吸なども取得できる。そこで、一般に生理心理学的指標を用いた情動分類手段では複数の指標を組合せることで精度が向上することが知られていることを踏まえ、これらの指標も加味して本方法の精度向上を図る。また、他の無拘束的なセンサの利用も検討する。例えば、ビデオを介した表情や顔色の取得などは効果的と考える。

- 本提案方法により得られる映像区間の認知的・情動的な特性を、映像作成の支援ツールに応用する。例えば、ラッシュフィルムの編集作業では、映像区間が与える視聴経験に関わる情報は有用であろう。また、試写会などで心拍活動データを得れば、作成した映像が意図した演出効果を挙げているかなども、ショットのようなきめ細かい単位で検証することができる。

- 選択した映像区間を単純に連結した要約映像は、評価実験で映像の構造や意味の観点から観づらいと評されたが、本研究の目標は区間選択方法の確立にあり、この問題は予期されたものである。しかし、実用上は重要な点であるので、適切な区間連結方法を検討していきたい。

- 本研究で得られた知見を用い、既存の映像要約技術との洗練された組合せ方法を確立する。例えば、1.3.3 節で説明した映像内情報を用いた経験型の映像要約方

法を前処理に用いて区間を粗く選択しておき、その上で本提案方法を利用するなどの方法が考えられる。

人間工学の分野では、映像視聴の経験を認知や情動の観点から分析する手法が多く確立されている。また、マルチメディアに関わる信号処理技術には、映像要約以外にも多様なものがある。そこで長期的な展望としては、上記に掲げた課題だけでなく、本研究の知見も含めて、人間工学的な手法をマルチメディア処理技術全般に幅広く導入していくことにより、メディアを利用するユーザを中心にした研究を展開していく予定である。

参考文献

[1] Aasman, J., Mulder, G. and Mulder, L.: “Operator Effort and the Measurement of Heart-Rate Variability”, Human Factors, **29**, 2, pp. 161-170 (1987)

[2] Affective Computing Group, MIT: `http://affect.media.mit.edu/`

[3] Aizawa, K., Ishijima, K. and Shiina, M.: “Automatic Summarization of Wearable Video”, Proc IEEE Pacific-Rim Conf Multimedia (PCM 2001), pp. 16-25 (2001)

[4] 相澤清晴, 石島健一郎, 椎名誠: “ウェアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み”, 信学論 D-II, **J86-D-II**, 6, pp. 807-815 (2003)

[5] Andreassi, J.: “Psychophysiology: Human Behavior & Physiological Response” (5th ed.), Lawrence Erlbaum, New Jersey (2007)

[6] Anttonen, J. and Surakka, V.: “Emotions and Heart Rate While Sitting on a Chair”, Proc ACM Conf Computer Human Interaction 2005, pp. 491-499 (2005)

[7] 青柳滋己, 藤孝治, 高田敏弘, 菅原俊治, 尾内理紀夫: “映像短縮再生システムの教育映像への適用評価”, 情処学論, **46**, 5, pp. 1297-1305 (2005)

[8] Arifin, S. and Cheung, P. Y. K.: “User Attention Based Arousal Content Modeling”, IEEE Int Conf Image Processing 2006 (ICIP 2006), pp. 433-436 (2006)

[9] 跡部裕貴, 泉正夫, 福永邦雄: “サッカーの放送型映像における試合中断区間に注目したイベント推定”, 信学技法, **PRMU2006-259**, pp. 25-30 (2007)

[10] Berntson, G. G., Cacioppo, J. T. and Fieldstone, A.: “Illusions, Arithmetic, and the Bidirectional Modulation of Vagal Control of the Heart”, Biological Psychology, **44**, 1, pp. 1-17 (1996)

[11] Berntson, G. G. and Cacioppo, J. T.: “Heart Rate Variability: Stress and Psychiatric Conditions”, in Malik, M. and Camm, A. J. (Eds.), “Dynamic Electrocardiography”, Futura, New York, pp. 57-64 (2004)

[12] Body Media, Inc., U.S.A.: `http://www.bodymedia.com/`

[13] Bordwell, D., Staiger, J. and Thompson, K.: “The Classical Hollywood Cinema: Film Style and Mode of Production to 1960”, Columbia University Press, New York (1985)

[14] Bordwell, D.: “The Way Hollywood Tells It: Story and Style in Modern Movies”, University of California Press, California (2006)

[15] Boreczky, J. S. and Rowe, L. A.: “Comparison of Video Shot Boundary Detection Techniques”, Proc SPIE Storage and Retrieval for Image and Video Database IV, pp. 170-179 (1996)

[16] Bradley, M. M.: “Emotion and Motivation”, in Cacioppo, J. T., Louis, L. and Berntson, G. G. (Eds.), “Handbook of Psychophysiology” (2nd ed.), Cambridge University Press, New York, pp. 602-642 (2000)

[17] Bradley, M. M., Codispoti, M., Cuthbert, B. N. and Lang, P. J.: “Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing”, Emotion, **1**, 3, pp. 276-298 (2001)

[18] チョコパラ TV, NTT：`http://www.chocopara.tv/`

[19] Czerwinski, M., Gage, D. W., Gemmell, J., Marshall, C. C., Pérez-Quiõnesis, Skeels, M. M. and Catarci, T.: “Digital Memories in an Era of Ubiquitous Computing and Abundant Storage”, CACM, **49**, 1, pp. 45-50 (2006)

[20] Dimitrova, N., Zhang, H., Shahraray, B., Sezan, I., Huang T. and Zakhor, A.: “Applications of Video-Content Analysis and Retrieval”, IEEE Multimedia, July-September 2002, pp. 42-55 (2002)

[21] Drescher, V. M., Grantt, W. H. and Whitehead, W. E.: “Heart Rate Response in Touch”, Psychosomatic Medicine, **42**, 6, pp. 559-565 (1980)

[22] 福田忠彦: “生体情報システム論”, 産業図書, 東京 (1995)

[23] 福井康之: “まなざしの心理学”, 創元社, 大阪 (1984)

[24] Hanjalic, A. and Xu, L. Q.: “User-Oriented Affective Video Content Analysis”, IEEE Workshop on Content-Based Access of Image and Video Libraries 2001 (CBAIVL 2001), pp. 50-57 (2001)

[25] 長谷山美紀, 久光徹: “情報大航海プロジェクトにおける共通技術”, 映情学誌, **63**, 1, pp. 42-47 (2009)

[26] 林良彦, 松永昭一, 松尾義博: “音声認識・言語処理の適用によるコンテンツ内容記述メタデータの生成”, NTT 技術ジャーナル, **15**, 4, pp. 21-24 (2003)

[27] Healey, J. and Picard, R.: “StartleCam: a Cybernetic Wearable Camera”, Proc Second Int Symposium on Wearable Computing, pp. 42-49 (1998)

[28] 日高浩太, 佐藤隆: “映像の即欄技術”, 映情学誌, **63**, 1, pp. 36-41 (2009)

[29] 本田麻子, 正木宏明, 山崎勝男: “情動喚起刺激に対する心臓血管系反応と脳波の偏側性”, 早大 人間科学研究, **15**, 1, pp. 39-45 (2002)

[30] 今井順一, 橋本誠, 金子正秀, 長島知正: “顔面の肌色解析による交感神経系活性度の非侵襲的評価”, 信学論 D, **J89-D**, 8, pp. 1869-1876 (2006)

[31] 入江豪, 日高浩太, 宮下直也, 佐藤隆, 谷口信行: “個人撮影映像を対象とした映像即覧のための “笑い” シーン検出法”，映情学誌，**62**，2, pp. 227-233 (2008)

[32] 石島健一郎, 椎名誠, 相澤清晴: “個人体験映像の構造化と要約 - 生体情報を用いた映像要約によるライフメディア”, 信学技報, **IE 2000-23**, pp. 51-58 (2000)

[33] 石坂和夫, 吉弘貢, 柳田昇: “映像信号記録装置，編集方法及びそのシステム”, 特開平 9-65271 (1997)

[34] 鎌田幹夫，坂東武彦，黒岩義之: “映像コンテンツ視聴時の生体信号計測・評価”, Proc 2001 JMAC Conf，pp. 11-1–11-4 (2001)

[35] Kang, H. B.: “Affective Content Detection using HMMs”, ACM Int Conf Multimedia 2003 (MM 2003)，pp. 259-262 (2003)

[36] Kapoor, A. and Picard, R. W.: “Multimodal Affect Recognition in Learning Environments”, ACM Int Conf Multimedia 2005 (MM 2005)，pp. 677-682 (2005)

[37] 加藤象二郎, 大久保尭夫 編: “初学者のための生体機能の測り方”, 日本出版サービス, 東京 (1999)

[38] 加藤和也, 吉高淳夫, 平川正人: “文脈を考慮に入れた映画の要約作成”, 情処学オーディオビジュアル複合処理研資, **AVM 36-5**, pp. 25-30 (2002)

[39] 小杉幸夫, 武者利光: “生体情報工学”, 森北出版, 東京 (2000)

[40] 経済産業省 情報大航海プロジェクト: `http://www.igvpj.jp/`

[41] Kerman, L.: “Coming Attractions - Reading American Movie Trailers”, University of Texas Press, Texas (2004)

[42] 財団法人 機械システム振興協会: “ストレス計測技術の安全対策への適用可能性”, 財団法人 人間生活工学研究センター報告書 (2004)

[43] 是津耕司, 上原邦昭, 田中克己: “映像の意味的構造”, 情処学論, **41**, 4, pp. 12-23 (2000)

[44] Lang, P. J., Bradley, M. M. and Cuthbert, B. N.: “International Affective Picture System (IAPS): Technical Manual and Affective Ratings”, University of Florida Technical Report, **A-6** (1997)

[45] Li, Y., Zhang, T. and Tretter, D.: “An Overview of Video Abstraction Techniques”, HP Technical Report, **HRL-2001-191** (2001)

[46] Leinhart, R.: “Comparison of Automatic Shot Boundary Detection Algorithms”, Proc SPIE Storage and Retrieval for Image and Video Database VII, pp. 290-301 (1999)

[47] 丸山隆, 廣井和重, 清水喜弘, 安達聡, 岩村真澄, 武井健, 四本直樹: “次世代 Prius を支える技術”, 日立評論, **87**, 8, pp. 35-38 (2005)

[48] 益満健, 越後富夫: “映像重要度を用いたパーソナライズ要約映像作成手法”, 信学論 D-II, **J84-D-II**, 8, pp. 1848-1855 (2001)

[49] 松山博輝, 佐々哲, 橘川千里: “映像コンテンツの編集装置及び編集方法”, 特開 2005–128884 (2005)

[50] MyLifeBits Project, Microsoft Research: `http://research.microsoft.com/barc/MediaPresence/MyLifeBits.aspx`

[51] 三浦宏一, 浜田玲子, 井出一郎, 坂井修一, 田中英彦: “動きに基づく料理映像の自動要約”, 情報学論 コンピュータビジョンとイメージメディア研究会, **44**, SIG 9, pp. 21-29 (2003)

[52] 宮森恒: “動作インデックスを用いた映像の自動注釈付けとその柔軟な内容検索への応用”, 情処学 情報学基礎研資, **2002**, 21 (FI-067-02), pp. 9-16 (2002)

[53] 宮田章裕, 福井健太郎, 本田研作, 重野寛, 岡田謙一: “会議を撮影した動画メディアの思考状態インデクシングの提案”, 情処学論, **45**, 11, pp. 2509-2518 (2004)

[54] 宮田章裕, 林剛史, 福井健太郎, 重野寛, 岡田謙一: “思考状態と発話停止点を利用した会議の動画ダイジェスト生成支援”, 情処学論, **47**, 3, pp. 906-914 (2006)

[55] Money, A. G. and Agius, H.: “Automating the Extraction of Emotion-Related Multimedia Semantics”, 19th British Computer Society Human Computer Interaction (BCS HCI) Group Annual Conf (2005)

[56] 森山剛, 坂内正夫: “ドラマ映像の心理的内容に基づいた要約映像の生成”, 信学論 D-II, **Vol. J84-D-II**, 6, pp. 1122-1131 (2001)

[57] 内閣府 経済社会総合研究所: “消費動向調査”, `http://www.esri.cao.go.jp/jp/stat/shouhi/shouhi.html`

[58] 中島義明: “映像の心理学”, サイエンス社, 東京 (1996)

[59] 中村裕一，外村佳伸: “見たい部分を簡単に短時間で”, 信学誌, **82**, 4, pp. 346-353 (1999)

[60] 中尾宗一郎: “ビデオカメラ”, 特開平 4–98979 (1992)

[61] NeuroSky, Inc., U.S.A.: `http://www.neurosky.com/`

[62] 新関久一, 高橋龍尚, 宮元嘉巳: “圧電ポリマー素子を用いたベッドサイド心拍・呼吸・体動モニタ”, 第 38 回日本日本エム・イー学会大会 (1998)

[63] 新関久一: “無拘束心拍・呼吸・体動エピソードの自動計測モニターの実用化”, 第 42 回日本日本エム・イー学会大会 (2003)

[64] 西田佳文, 武田正資, 森武俊, 溝口博, 佐藤知正: “圧力センサによる睡眠中の呼吸・体位の無侵襲・無拘束な計測”, 日本ロボット学会誌, **16**, 5, pp. 705-711 (1998)

[65] NTT 技術ジャーナル編集部: “コンテンツ時短視聴技術 Chocopara について教えてください”, NTT 技術ジャーナル, **17**, 5, pp. 88-89 (2005)

[66] Öhman, A., Hamm, A. and Hugdahl, K.: “Cognition and the Automatic Nervous System”, in Cacioppo, J. T., Louis, L. and Berntson, G. G. (Eds.), “Handbook of Psychophysiology” (2nd ed.), Cambridge University Press, New York, pp. 533-575 (2000)

[67] Picard, R. W.: “Affective Computing”, The MIT Press, Massachusetts (1997)

[68] Picard, R. W., Vyzas, E. and Healey, J.: “Toward Machine Emotional Intelligence: Analysis of Affective Physiological State”, IEEE Trans Pattern Anal. & Mach. Intell., **23**, 10, pp. 1175-1191 (2001)

[69] Polar Electro, Finland: `http://www.polar.jp/`

[70] Ritvanen, T., Latinen, T. and Hänninen, O.: “Relief of Work Stress after Weekend and Holiday Season in High School Teachers”, J Occup Health, **46**, 3, pp. 213-215 (2004)

[71] Rowe, D., Sibert, J. and Irwin, D.: “Heart Rate Variability: Indicator of User State as an Aid to Human-Coputer Interaction”, Proc ACM Conf Computer Human Interaction 1998, pp. 480-487 (1998)

[72] 佐藤美穂, 間峠慎吾, 森俊夫, 鈴木昇, 春日正男: “体感センサを利用した映像コンテンツの印象計測の検討”, 映情学誌, **60**, 4, pp. 425-430 (2006)

[73] Shibata, Y., Takahashi, Y., Kamada, M., Osawa, E., Kimura, H. and Miura, M.: “Weight DS Extension for Affect-Based Content Characterization”, ISO/IEC JTC1/SC29/WG11, MPEG 88/M5481 (1999)

[74] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. and Jain, R.: “Content-Based Image Retrieval at the End of the Early Years”, IEEE Trans Pattern Anal. & Mach. Intell., **22**, 12, pp. 1348-1380 (2000)

[75] 総務省 情報通信政策研究所: “メディア・ソフトの製作及び流通の実体”, `http://www.soumu.go.jp/s-news/2008/080702_1.html` (2008)

[76] Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G. and Picard, R. W.: “The HandWave Bluetooth Skin Conductance Sensor”, Int Conf Affective Computing and Intelligent Interaction, pp. 699-706 (2005)

[77] 杉田典大, 吉澤誠, 田中明, 阿部健一, 山家智之, 仁田新一: “血圧-心拍間の最大相互相関係数を用いた映像刺激の生体影響評価”, ヒューマンインタフェース学会論文誌, **4**, 4, pp. 227-234 (2002)

[78] 鈴木哲, 松井岳巳, 藤江翔吾: “マイクロ波レーダーを用いた作業中における心拍変動指標の非接触計測の試み”, 人間工学, **48**, Supplement, pp. 272-273 (2008)

[79] 高橋和彦: “マルチモーダル生体信号情報による感情認識に関する一考察”, 人間工学会誌, **41**, 4, pp. 248-253 (2005)

[80] 滝嶋康弘: “映像の自動要約技術”, 映情学誌, **62**, 5, pp. 714-716 (2008)

[81] 谷口行信, 外村佳伸, 浜田洋: “映像ショット切換え検出法とその映像アクセスインタフェースへの応用”, 信学論 D-II, **J79-D-II**, 4, pp. 538-546 (1996)

[82] Task Force of The European Society of Cardiology & The Northern American Society of Pacing and Electrophysiology: “Heart Rate Variability: Standards of Measurements, Physiological Interpretation, and Clinical Use”, European Heart Journal, **17**, pp. 354-381 (1996)

[83] Teeters, A., Kaliouby, R. El and Picard, R. W.: “Self-Cam: Feedback from What Would be Your Social Partner”, Proc ACM SIGGPRAH 2006, p. 138 (2006)

[84] TRECVID (Text Retrieval Conference - Video Retrieval Evaluation), NIST, U.S.A.: http://www-nlpir.nist.gov/projects/trecvid/

[85] Truong, B. T. and Venkatesh, S.: "A Video Abstraction: A Systematic Review and Classification", ACM Trans Multimedia Computing, Communications & Applications, **3**, 1, Article 3 (2007).

[86] Xu, M., Chia, L. T. and Jin, J.: "Affective Content Analysis in Comedy and Horror Videos by Audio Emotional Event Detection", IEEE Int Conf Multimedia and Expo 2005, pp. 622-625 (2005)

[87] 山中幸治, 田島隆行, 小栗宏次, 岩田彰: "感圧センサを用いた呼吸成分抽出アルゴリズムの検討", 第 22 回医療情報学連合大会 (JCMI 2002), pp. 244-245 (2002)

[88] 吉野公三: "不安全行動の計測と理解", ヒューマンインタフェース学会誌, **8**, 3, pp. 179-182 (2006)

[89] 吉高淳夫, 松井亮治, 平嶋宗: "カメラワークを利用した感性情報の抽出", 情処学論, **47**, 6, pp. 1696-1707 (2006)

謝辞

最初に、時には仕事に、時には家事と育児に追われて研究活動をなおざりにしてきた私を、ここまで見放さずに研究指導をして下さった早稲田大学大学院 国際情報通信研究科（GITS）の河合隆史教授に感謝したい。河合研究室の皆様、特に機材や計算機環境の細々した手配を手伝ってくれた盛川浩志君、統計処理の細かい点をご教授して下さった国際情報通信センターの三家礼子客員准教授、この論文に丁寧なチェックを入れてくれた太田啓路君にはお世話になった。岸君、Kim 君、加藤君、阿部君をはじめとする諸君には、トシだけは上なのを嵩に着て、忙しい中を被験者を強要してきてしまった。諸兄には、協力感謝と手を合せると共に、怪しい中年男と遊んでくれたことに感謝したい。また、研究活動を再開する前からいろいろと相談に乗ってくれた GITS の渡辺裕教授、本論文作成にあたって貴重な意見を下さった GITS の亀山渉教授と坂井滋和教授、作成した要約映像にコメント下さった早稲田大学芸術学校の藪野健教授にも心からの謝意を示したい。

また、研究家業の楽しさを教えてくれた母校 国際基督教大学 物理学教室の諸先生方、NTT ヒューマンインタフェース研究所時代の友人や諸先輩らにも感謝したい。

6 年間に及ぶ「趣味」の研究生活は、仕事を廻してくれた人たちがいなければ経済的には支えきれなかった。特に、カットシステムの石塚勝敏社長、アスキーメディアワークス ネットワークマガジン編集部の武堂貴宏氏（現 F5 ネットワークスジャパン）、石山俊浩氏（現 ドット PC 編集部）、臼田良寛氏、大島伸一編集長（元）、及び大谷イビサ編集長（現）、オライリージャパンの宮川直樹氏に感謝したい。また、モノ書き家業を最初に紹介してくれた森崎さんには感謝の言葉もない。副業の目処がたっていなかったら、大学に行こうとは思わなかっただろう。高時給で結婚・出産費用を一気に調達させてくれたシマンテック（旧 Veritas Software Corporation）の Honma さんと

Sakai さんにも一言感謝の言葉を述べたい。

最後に、亡き父に、フリーターともつかない愚息に暖かい支援を送ってくれた母に、応援してくれた姉に、稼ぎもロクにない主夫に耐えてくれた妻に、Joana と Elizabeth に、そして無限の愛と喜びを与えてくれる茉莉枝に本論文を捧げたい。

February 2009

Satoshi Toyosawa

研究業績

論文誌論文

1. 豊沢聡, 河合隆史: “心拍変動を利用した短縮映像作成方法”, ヒューマンインタフェース学会論文誌, **9**, 2, pp. 173-179 (May 2007)

2. 豊沢聡, 河合隆史: “視聴者の心拍活動を用いた映像短縮方法とその評価”, 映情学誌, **63**, 1, pp. 86-94 (January 2009)

国際会議

1. Toyosawa, S. and Kawai, T.: “Video Digest Based on Heart Rate”, 7th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2007), pp. 15-20 (August 2007)

2. Toyosawa, S. and Kawai, T.: “An Assemblage of Impressive Shots - a Video Digesting Method Based on Viewer’s Heart Activity”, 12th IASTED International Conference on Internet and Multimedia Systems and Applications (IMSA 2008), pp. 107-112 (August, 2008)

学会発表

1. 豊沢聡, 河合隆史: “心拍変動を利用した短縮映像の物理的特性について”, 人間工学, **42**, Supplement, pp. 206-207 (June, 2006)

2. 豊沢聡, 河合隆史: “心拍変動を利用した短縮映像の主観的評価”, 2006 年映情学

年次大, Session 19-5, (August 2006)

3. 豊沢聡, 河合隆史: “心拍動を利用した個人的な印象メモとしての短縮映像”, 人間工学会関東支部第 37 回大会, pp. 89-90 (November 2007)

4. 豊沢聡, 河合隆史: “視聴者の注意指標に基づいた短縮映像の生成方法”, 人間工学, **44**, Supplement, pp. 298-299 (June 2008)

その他の業績

1. 豊沢聡, 河合隆史: “画像中の色数を用いた空間のヒト混雑度判定”, 人間工学, **40**, Supplement, pp. 448-449 (June 2004)

2. 豊沢聡, 河合隆史: “フレーム間差分を用いた空間の混雑度判定”, 2004 年映情学年次大, Session 3-8 (August 2004)

3. 豊沢聡, 河合隆史: “静止画像の色彩から得る空間の混雑度状況”, ヒューマンインタフェースシンポジウム 2004, pp. 555-560 (October 2004)

4. 豊沢聡, 河合隆史: “歩行中の滞留のし易さを考慮した歩行空間の混雑度判定”, 第 67 回情処学全大, **4**, pp.15-16 (March 2005)

5. Toyosawa, S. and Kawai, T.: “Crowdedness Evaluation in Public Space for Pedestrian Guidance”, 8th International IEEE Conference on Intelligent Transportation Systems (ITSC 2005), pp. 137-142 (September 2005)

6. 豊沢聡: “実践 Java ネットワークプログラミング”, カットシステム, 東京 (July 2002)

7. 豊沢聡訳, 金崎裕己監訳: “SAN & NAS ストレージエリアネットワーク”, オライリージャパン, 東京 (October 2002)

8. 豊沢聡訳, 市原英也監訳: “IPv6 エッセンシャルズ”, オライリージャパン, 東京 (April 2003)

9. 豊沢聡訳, 市原英也監訳: “IPv6 エッセンシャルズ 第 2 版”, オライリージャパン, 東京 (June 2007)

10. 豊沢聡, のりぞう: “コマンドで理解する TCP/IP”, アスキー, 東京 (March 2008)